

Saidha, Tekchand

From: Saidha, Tekchand
Sent: Monday, June 23, 2003 2:47 PM
To: STIC-Biotech/ChemLib
Subject: sequence search request - 09/830751

09/830751

Please search the data base and interference files for :

SEQ ID Nos. 2, 4, 6 and 8

Thank you !

Tekchand Saidha

Primary Examiner

Art Unit 1652, CM1, Room No. 10D05

Mail Box 10D01

(703) 305-6595

The first part of the paper discusses the importance of understanding the cultural context of the research. It highlights the need for researchers to be sensitive to the values and beliefs of the communities they are studying. This is particularly important in the field of education, where cultural differences can significantly impact learning outcomes.

The second part of the paper focuses on the methodology used in the study. It describes the process of selecting participants, collecting data, and analyzing the results. The authors emphasize the importance of using a mixed-methods approach to gain a comprehensive understanding of the research topic.

The third part of the paper presents the findings of the study. It discusses the results of the quantitative data analysis and the insights gained from the qualitative interviews. The authors conclude that there are significant differences in learning outcomes between the two groups, and these differences can be attributed to cultural factors.

The final part of the paper offers recommendations for future research and practice. It suggests that educators should be aware of the cultural context of their students and tailor their teaching methods accordingly. Additionally, it calls for further research to explore the underlying reasons for the observed differences.



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number

TO: Tekchand Saidha
Location: CM1/10D05/10D01
Art Unit: 1652
Tuesday, June 24, 2003

Case Serial Number: 830751

From: Edward Hart
Location: Biotech-Chem Library
CM1-6B02
Phone: 305-9203

edward.hart@uspto.gov

Search Notes

Examiner Saidha,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart



97247

STIC-Biotech/ChemLib

From: Saidha, Tekchand
Sent: Monday, June 23, 2003 2:47 PM
To: STIC-Biotech/ChemLib
Subject: sequence search request - 09/830751

09/830751

Please search the data base and interference files for :

SEQ ID Nos. 2, 4, 6 and 8

Thank you !

Jekchand Saidha
Primary Examiner
Art Unit 1652, CM1, Room No. 10D05
Mail Box 10D01
(703) 305-6595

CRIFE

6/24/03
17B5502
4-ATT

100
100



STIC SEARCH RESULTS

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact:*

Mary Hale, Information Branch Supervisor
308-4258, CM1-1E01

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup: Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library CM1-Circ Desk



[The body of the document contains extremely faint, illegible text, likely bleed-through from the reverse side of the page. The text is organized into several paragraphs, but the specific content cannot be discerned.]

GenCore version 5.1.6
Copyright (c) 1993 2003 Compugen Ltd.

CM protein - protein search, using sw model

Run on: June 24, 2003, 10:02:25 Search time 28.8822 seconds

(without alignments)
1668483 Million cell updates/sec

Title: us-09-830-751-2

Perfect score: 2560
Sequence: 1 MSHLPMTVPIKLPMNTEYEQ MSVIALCNYLQKAVAKILD 445

Scoring table:
Gap: 10.0, Gap ext: 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying above parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

A_Geneseq.101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2560	100.0	495	2	AAH74923
2	1405.5	55.1	497	2	AAH60456
3	1399	54.9	496	16	AAH71891
4	1362.5	53.4	495	16	AAH71803
5	1248.5	49.0	517	23	AAH44516
6	1235.5	48.6	517	23	AAH49517
7	1225.5	48.1	500	22	AAH74924
8	1218.5	47.8	512	23	AAH61842
9	1218.5	47.8	512	23	AAH73864
10	1218.5	47.8	529	22	AAH00577

11	1203.5	47.2	520	22	AAH60140
12	1201	47.1	538	21	AAH63412
13	1179.5	46.2	559	22	AAH62252
14	1169.5	45.9	534	21	AAH67414
15	1153.5	45.2	521	15	AAH64672
16	1146.5	45.0	501	21	AAH60439
17	1146.5	45.0	501	21	AAH67413
18	1073	42.1	923	23	AAH68241
19	1047	40.7	909	22	AAH64519
20	1045	40.7	906	22	AAH74957
21	1020.5	40.0	424	21	AAH36240
22	1015	39.8	508	18	AAH89293
23	982.5	38.5	510	22	AAH36695
24	972.5	38.1	477	22	AAH34883
25	970.5	38.1	496	22	AAH34148
26	970.5	38.1	496	22	AAH37064
27	949.5	37.2	496	22	AAH81968
28	949.5	37.2	506	23	AAH39538
29	941.5	36.9	457	21	AAH36241
30	940	36.8	490	22	AAH36509
31	945	36.7	512	22	AAH64536
32	921.5	36.1	485	23	AAH05196
33	921.5	36.1	485	23	AAH07141
34	921.5	36.1	508	21	AAH00082
35	921.5	36.1	508	23	AAH16744
36	919.5	36.1	506	22	AAH49208
37	919	36.0	490	22	AAH34464
38	895	35.1	506	22	AAH92818
39	895	35.1	506	22	AAH79351
40	890.5	34.9	507	22	AAH80035
41	890.5	34.9	507	22	AAH80043
42	888	34.8	497	15	AAH63766
43	882	34.6	501	21	AAH10881
44	881.5	34.6	497	21	AAH10882
45	881.5	34.6	503	21	AAH35662

ALIGNMENTS

RESULT 1
ID: AAH74923 standard. Predicted 495 AA.
XX
AC: AAH74923;
XX
DT: 26-JUN-2001 (first entry)
XX
DE: yeast ald-hyde dehydrogenase. Ald4 protein sequence sbg 11. MW:2.
XX
KW: Ald-hyde dehydrogenase; glycol; dehydratase; 3-mer; glycerol;
KW: leucyl-oxo; 3-hydroxypropionic acid; genetic engineering; glucose;
KW: lactate; host; absorbable plastic; device; surgical; suture;
KW: beta-lactam; acrylic acid; trifluoromethylated alcohol; diol;
KW: polyhydroxyalkonate; copolymer; lactic acid.
XX
OS: Saccharomyces cerevisiae.
XX
PN: W0200116346-A1.
XX
PD: 08-MAR-2001.
XX
PF: 30-AUG-2000; 2000W-US248/8.
XX
PP: 30-AUG-1999; 99US-0151440.
XX
PA: (WISC) WISCONSIN ALUMNI PRESS FOUND.
XX
PI: Suthers PF, Cameron DC.
XX
DR: WPI: 2001-315988/33.
XX
DR: N-FSDB: AAH82081.
XX

4 Hydroxyprotonic acid preparation, for use e.g. as monomer, by fermenting recombinant microorganisms expressing genes for suitable enzymes in the presence of glycerol or glucose -
 claim 5, Page 26-28, 64pp, English.

The present invention describes a method for the production of 3 hydroxyprotonic acid (3-HP). The method comprises fermenting a recombinant microorganism in the presence of a source of glycerol (1) or glucose, where the microorganism: (i) expresses genes for non-native enzymes which catalyse the production of (3-HP) from (1); (ii) carries genetic constructions for the expression of a glycerol dehydratase (GDH) and aldehyde dehydrogenase (ADH) capable of catalysing the production of (3-HP) from (1); or (iii) carries a genetic construct which expresses the dhah gene from *Klebsiella pneumoniae* and a gene for an ADH capable of catalysing the production of (3-HP) from (1). 3-HP is a monomer, and is useful e.g. in the production of absorbable prosthetic devices and surgical sutures or for incorporation into beta-lactams, production of acyclic acid or formation of trifluoromethylated alcohols or diols, polyhydroxyalkonates and copolymers with lactic acid. Incorporation of genes encoding two enzymes makes the host organisms able to produce (3-HP) from (1). The biotechnological method of preparing (3-HP) is potentially cheaper than chemical synthesis. The present sequence represents the yeast aldehyde dehydrogenase AldA, which is used in the exemplification of the present invention.

Sequence 495 AA:

Query Match 100.0%; Score 2550; DB 22; Length 495;
 Best Local Similarity 100.0%; Pred. No. 1.7e-207;
 Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

UY 1 MSHLPMTVPIKLNGLDEYEGTLEINNKFPVSKUNKFEVFNPSIEELCHLYEGREDD 60
DB 1 MSHLPMTVPIKLNGLDEYEGTLEINNKFPVSKUNKFEVFNPSIEELCHLYEGREDD 60
UY 61 VEFVAVQADAPAFNSGNSNNIPIDPDKKALYRIAEHLEQDKVDIASLETIDNKAISSSG 120
DB 61 VEFVAVQADAPAFNSGNSNNIPIDPDKKALYRIAEHLEQDKVDIASLETIDNKAISSSG 120
UY 121 DMDLVINYLKSSAGFAFKDKRMIDTGRTHSYTKKQPLGVCGLIPMNPPLIMWAKTA 180
DB 121 DMDLVINYLKSSAGFAFKDKRMIDTGRTHSYTKKQPLGVCGLIPMNPPLIMWAKTA 180
UY 181 PALVTGNTVYIKTAESTFHSATYVSKYTPQAGIPPCVINIVSGEGKIVEAATINPKIKK 240
DB 181 PALVTGNTVYIKTAESTFHSATYVSKYTPQAGIPPCVINIVSGEGKIVEAATINPKIKK 240
UY 241 VAFVGSATATGRHITYSAAAGIKKVTLELGGKSNPIYFADAEIKKAVQNTILGITYNSGEV 300
DB 241 VAFVGSATATGRHITYSAAAGIKKVTLELGGKSNPIYFADAEIKKAVQNTILGITYNSGEV 300
UY 401 CTAGSPVYVEESTYDKFIEEFKAASESKVGPFOESTPGCAOTSOMQINKILKYVDIK 360
DB 401 CTAGSPVYVEESTYDKFIEEFKAASESKVGPFOESTPGCAOTSOMQINKILKYVDIK 360
UY 401 CTAGSPVYVEESTYDKFIEEFKAASESKVGPFOESTPGCAOTSOMQINKILKYVDIK 360
DB 401 CTAGSPVYVEESTYDKFIEEFKAASESKVGPFOESTPGCAOTSOMQINKILKYVDIK 360
UY 461 NEZATLITGGERIGSKSYFKPTVPGIVKEMKIVKEELIPGVVTVTKRSADAEVINMAN 420
DB 461 NEZATLITGGERIGSKSYFKPTVPGIVKEMKIVKEELIPGVVTVTKRSADAEVINMAN 420
UY 421 DSEFGIAGAGIHISNTINAIKVAUVNANGLVNTNLYPDMAVPEGSPNNSGLSPMSVIA 480
DB 421 DSEFGIAGAGIHISNTINAIKVAUVNANGLVNTNLYPDMAVPEGSPNNSGLSPMSVIA 480
UY 481 LGNTLVKKAVAKID 495
DB 481 LGNTLVKKAVAKID 495

```

RESULT 2
 AAF60456 standard: Proteins 497 AA.
 DB AAF60456
 XX

AC AAF60456.
 XX 25-JUN-1991 (first entry)
 DT

XX Sequence of aldehyde dehydrogenase (aldA) of *Aspergillus nidulans*.

XX Filamentous fungi promoter.

XX *Aspergillus nidulans*.

XX W08506097-A.

XX 23-OCT-1986.

XX 14-APR-1986; 86W0-GB00209.

XX 20-DEC-1985; 85US-0811404.

XX 15-APR-1985; 85CA-0479135.

XX (ALF-) ALLELIX INC.

XX Gwynne D.I., Buxton F., Pickett M., Davies R., Scazzocchio C;

XX WPI: 1986-291664/44.

XX N-PSDB: AAN60401.

XX DNA construct for use in filamentous fungi - comprising promoter

XX operative in filamentous fungi to promote transcription of coding

XX region

XX Disclosure: Fig 1A; 75pp; English.

In the constructs of the invention, the promoter region naturally associated with the alcohol dehydrogenase I (aldA) gene and the aldehyde dehydrogenase (aldA) gene of *A. nidulans* or naturally associated with the glucanase gene in *Aspergillus niger* may be used. The DNA construct may contain a promoter region in operative association with a signal peptide coding region. The promoter/signal construct is suitably provided with a flanking restriction site to allow precise coupling of the protein coding region to the signal peptide coding region.

Sequence 497 AA:

Query Match 55.1%; Score 1405.5; DB 7; Length 497;
 Best Local Similarity 54.3%; Pred. No. 1.9e-110;
 Matches 269; Conservative 88; Mismatches 133; Indels 5; Gaps 3;

```

UY 1 MSHLPMTVPIKLNGLDEYEGTLEINNKFPVSKUNKFEVFNPSIEELCHLYEGREDD 60
DB 1 MSHLPMTVPIKLNGLDEYEGTLEINNKFPVSKUNKFEVFNPSIEELCHLYEGREDD 60
UY 61 VEFVAVQADAPAFNSGNSNNIPIDPDKKALYRIAEHLEQDKVDIASLETIDNKAISSSG 120
DB 61 VEFVAVQADAPAFNSGNSNNIPIDPDKKALYRIAEHLEQDKVDIASLETIDNKAISSSG 120
UY 121 DMDLVINYLKSSAGFAFKDKRMIDTGRTHSYTKKQPLGVCGLIPMNPPLIMWAKTA 180
DB 121 DMDLVINYLKSSAGFAFKDKRMIDTGRTHSYTKKQPLGVCGLIPMNPPLIMWAKTA 180
UY 181 PALVTGNTVYIKTAESTFHSATYVSKYTPQAGIPPCVINIVSGEGKIVEAATINPKIKK 240
DB 181 PALVTGNTVYIKTAESTFHSATYVSKYTPQAGIPPCVINIVSGEGKIVEAATINPKIKK 240
UY 241 VAFVGSATATGRHITYSAAAGIKKVTLELGGKSNPIYFADAEIKKAVQNTILGITYNSGEV 300
DB 241 VAFVGSATATGRHITYSAAAGIKKVTLELGGKSNPIYFADAEIKKAVQNTILGITYNSGEV 300
UY 401 CTAGSPVYVEESTYDKFIEEFKAASESKVGPFOESTPGCAOTSOMQINKILKYVDIK 360
DB 401 CTAGSPVYVEESTYDKFIEEFKAASESKVGPFOESTPGCAOTSOMQINKILKYVDIK 360
UY 401 CTAGSPVYVEESTYDKFIEEFKAASESKVGPFOESTPGCAOTSOMQINKILKYVDIK 360
DB 401 CTAGSPVYVEESTYDKFIEEFKAASESKVGPFOESTPGCAOTSOMQINKILKYVDIK 360
UY 461 NEZATLITGGERIGSKSYFKPTVPGIVKEMKIVKEELIPGVVTVTKRSADAEVINMAN 420
DB 461 NEZATLITGGERIGSKSYFKPTVPGIVKEMKIVKEELIPGVVTVTKRSADAEVINMAN 420
UY 421 DSEFGIAGAGIHISNTINAIKVAUVNANGLVNTNLYPDMAVPEGSPNNSGLSPMSVIA 480
DB 421 DSEFGIAGAGIHISNTINAIKVAUVNANGLVNTNLYPDMAVPEGSPNNSGLSPMSVIA 480
UY 481 LGNTLVKKAVAKID 497
DB 481 LGNTLVKKAVAKID 497

```


Db 357 KKAAGTAVATGGDREHENEVTPQPPVTVTVTSDMKTAQEFIPGPVATQKREDAFAKRTG 416
 420 NSESEGLAAGHTSNTINALKVADRVNAGTWININVDHFAVPFGSEFNASGLSKREMSVD 479
 417 XSTQGLAAAVTRKVNINAIKVSNAKAGIYWINNNKINISVQAFPGHFKSGSLCPREISY 476
 QY 480 ALONTLOKAVAKL 494
 477 ALENTYQIKIVHYRL 491

RESULT 3

AAK71891
 ID AAK71891 standard. Protein: 496 AA.

AC AAK71891:
 DT 25-OCT-1995 (first entry)

DE Cladosporium herbarum allergen ClaH53.

KW Fungal spore, allergen, ClaH53, allergy: aldehyde dehydrogenase.

OS Cladosporium herbarum

PN W09506121-A.

PD 02-MAR-1995.

PF 24-AUG-1994: 94W0-AT90120.

PR 27-AUG-1993: 93AT-0001725.

PA (BIOM-) BIOMAY PRODN & HANDELSGES MBH.

PI Achatz G., Breitenbach M., Ebner C., Hirschwehr R.

PI Kraft D., Lechenauer E., Oberkofler H., Prillinger H.

PI Simon B., Unger A.

DR WPI: 1995-106850/14.

DR N-PSDB: AA086278.

PS Claim 1, Page 8 3, 35pp; German.

CC Spores of Cladosporium herbarum are the most common fungal spores

CC found in the air; they can cause allergic reactions. Various ClaH

CC allergens and sequences encoding them have now been isolated. The

CC mature ClaH53 allergen has mol. wt. 53 kD and is encoded by cDNA

CC sequence AA086278. The allergen has homology to aldehyde

CC dehydrogenases. Potential epitopic subfragments were identified by

CC computer analysis of the amino acid sequence. See AAK71892-471906 for

CC potential B-cell epitopes and AAK72615-472627 for potential T-cell

CC epitopes.

XX Sequence 496 AA;

SO

Query Match 54.9%; Score 1399, DR 16, Length 496;

Best Local Similarity 53.8%; Pred. No. 6, 7e-110;

Matches 263; Conservative 89; Mismatches 135; Indels 2; Gaps 2.

QY 7 TVPIKIPNGTKEPQPTGIFITNNKVPVSKONTFPVINDSTFEFICHTYEREDDVEFAVD 66

DB 3 SVETLPHSKYKTEPTGLFINNEFYKGECKTFEIVINSDESVILGVHEAIEKLVDAVA 62

QY 67 AADRFAFNSNMGITPTTFCKATVYAEIIFGKIVIASITLNNKAISSKQGVAVLI 126

DB 63 AAKQAF-FTSWPLETFPNNKRLNNIANKFNITDILAAVSSILNNKAISSMAVTSN'AS 121

QY 127 NTKSSAPFARKITGGMTTIGTHSTYKQPLGVWQQLIPWNPILIMAKITAPALVTS 186
 122 GFLPYVYQWALKITISKVITITPTTFENVYKKEPLISVPSCHSLPTPLIMWAKIGFALAG 181
 QY 187 NTVEKIKESPLSALYSEKIPQATPEVINTVSSFEKIVEEATINHKIKVAFETS 246
 182 NTIVLKTAMQPTPLGLVAASIVKAGPFGVINYISGFGVACAAALSSHDVQKVAFTOS 243
 QY 247 TATPEHYQSA-AATKKVTLFTGSEKSPNIVFAVAFELKRAVNTITGVYNSSTEVTAQS 305
 242 TVVSGTILKAASSNLKAVTELEDSKSPNIVFAVAFELKRAVNTITGVYNSSTEVTAQS 301
 QY 306 PVYVPESTYVQKTEPKKASNSIKVQPPRISTVYQAGTQMDNKTIKVQIDKQKQAT 365
 302 PVYVPESTYVQKTEPKKASNSIKVQPPRISTVYQAGTQMDNKTIKVQIDKQKQAT 361
 QY 366 TTTGGEPISGKGYPTKPPVPGVQKEDMPYVEETFEVAVITFEKSADEVINANNSEFG 425
 362 VETGSEKQDKQYLETETFEVQEDMKIVKELTFEGVCSIAFTPTTATATIGTAATY 421
 DB 422 LAAAVHTKNLNIALLVSNALKAQIWMVYVNLHHCMPFGYKESQIGREIAGDALANT 481
 QY 486 CYKAVPAKL 494
 DB 482 QRTKVSIRL 490

RESULT 4

AAK71803
 ID AAK71803 standard. Protein: 495 AA.

AC AAK71803:

DT 02-NOV-1995 (first entry)

DE Alternaria alternata allergen Alta53.

KW Fungal spore, allergen, Alta53, allergy: aldehyde dehydrogenase.

OS Alternaria alternata.

PN W09506122-A.

PD 02-MAR-1995.

PF 24-AUG-1994: 94W0-AT90121.

PR 27-AUG-1993: 93AT-0001725.

PA (BIOM-) BIOMAY PRODN & HANDELSGES MBH.

PI Achatz G., Breitenbach M., Ebner C., Hirschwehr R.

PI Kraft D., Lechenauer E., Oberkofler H., Prillinger H.

PI Simon B., Unger A.

DR WPI: 1995-106851/14.

DR N-PSDB: AA086275.

PS Claim 7, Page 8; 26pp; German.

CC Allergens derived from Alternaria alternata - their isolation by

CC reacting sera from patients with fungal extracts; useful for

CC in-vitro allergy detection.

XX Species of Alternaria alternata can cause allergic reactions. Various

XX AltA allergens and sequences encoding them have now been isolated.

XX The mature Alta53 allergen has mol. wt. 53 kD and is encoded by cDNA

XX sequence AA086275. The allergen has homology to aldehyde

XX dehydrogenases. Potential epitopic subfragments were identified by

XX computer analysis of the amino acid sequence. See AAK71804-471816 for

XX potential B-cell epitopes and AAK71817-471832 for potential T-cell

XX	RESULT 6
XX	AA049517
XX	10 AA049517 standard; Protein: 517 AA
XX	AA049517:
XX	13-MAY-2002 (first entry)
DE	Human ALDH5 protein polymorphic variant
XX	ALDH5, human, polymorphism, haplotype, aldehyde dehydrogenase 5,
KW	binding affinity; drug targeting, alcoholism, alcohol induced disorder,
KW	antialcoholic.
XX	
XX	Human sapiens
XX	
XX	Key Location/Qualifiers
XX	
XX	Misc-difference 86
XX	/label= Ala, Val
XX	
XX	Misc-difference 107
XX	/label= Leu, Arg
XX	
XX	Misc-difference 176
XX	/label= Val, Ile
XX	
XX	Misc-difference 253
XX	/label= Met, Val
XX	
XX	Misc-difference 276
XX	/label= Asp, Asn
XX	
XX	Misc-difference 470
XX	/label= Val, Ala
XX	
XX	WO/00192379 A2
XX	
XX	06-DEC-2001
XX	
XX	24-MAY-2001: 2001WO-0517253
XX	
XX	24-MAY-2000: 2000GNS 267508F
XX	
XX	(GFWA-) GENMISANAT PHARM INC.
XX	
XX	Duda A, Finkel K, Kazemi A, Messer C, Sanchis A;
XX	WPI: 2002:122054/16.
XX	
XX	New genetic variants with polymorphisms in the aldehyde dehydrogenase 5
XX	(ALDH5) gene, useful for studying the function of ALDH5, and for
XX	expressing ALDH5 protein which is useful in screening drugs for
XX	treating ALDH5-related diseases
XX	
XX	Disclosure: Fig 3; 9ppp; English.
XX	
XX	This invention describes a novel isolated genes and haplotypes of the
XX	human aldehyde dehydrogenase 5 (ALDH5) gene containing polymorphic sites
XX	The polymorphic ALDH5 variant is useful in studying the effect of the
XX	variation on the biological activity of ALDH5 and on the binding
XX	affinity of candidate drugs targeting ALDH5 for the treatment of
XX	alcoholism and alcohol-induced disorders. Polymorphisms comprising a
XX	polymorphic gene variant or fragment may be used for therapeutic
XX	purposes. ALDH5 protein isoforms may be used in assays to measure the
XX	binding affinities of one or more candidate drugs targeting the ALDH5
XX	protein. ALDH5 proteins may be used to generate antibodies. Haplotyping
XX	method can be used by scientists to validate ALDH5 as a candidate
XX	target for treating a specific condition or disease predicted to be
XX	associated with ALDH5 activity, and in the design of clinical trials of
XX	candidate drugs for treating a specific condition or disease predicted
XX	to be associated with ALDH5 activity. Information on polymorphisms or
XX	the ALDH5 gene can be applied for studying the biological function of
XX	ALDH5 as well as in identifying drugs targeting this protein for the
XX	treatment of disorders related to its abnormal expression or function.
XX	The products of the invention have antialcoholic activity. This sequence
XX	represents the human ALDH5 polymorphic variant protein described in the

QY	21	DFID65KALYPDLAEITFGRKIVFASIFETLNTKRA-TSSSRGIVWIVAVINYLKSSAGFAIKLI	139
Db	92	UALS5PDLHSLAULVERKWAIAAETMTGKRPFFIHAPFIDIEKIPITPLPYAFCAWAKLI	151
QY	140	IGCKMIDITKTHFSTKPGKPGVGVQGLIPWNPFPILMMAWKLAVALVGNINIVVLKTAESFPL	199
Db	152	QKTKTIPITGNNVPTFRHEPFSVQYATIPWNPFLMLVLVWKIAPALQNSNMLVLRPAEQIPL	213
QY	200	SALVYSKYIPQAGTIPPGAVINIVSGPSKIVVFATITNHPKIKKYAFTGSTATGPHVQSAA-	258
Db	212	YALYGLSLIKKAGFPFGVWIVFCGPTVCAALSSHPQINKIAFTGSTEVGKLVKEAASR	271
QY	259	AGLKVKVTELEGGKSPNIVFADAEIKKAVQNIITICTVYNSGPGVAGSPVYVVESTYDKPT	318
Db	272	SNLKVATELEGGKKNCLWALADGLDAVEAHQGVFENQVQNTAAASPVFVEEQVYSEFV	331
QY	319	EEPRAAASIKVGHPHSTSPFGCAQTSMQMLNKILKAYVDLGNKGATILITGGERLGSKEY	378
Db	332	RRSEYAKRRVPGSPFQVKEQGPILQKGFUKILEITPSGKKKSAKIPFGSSAMFDKSL	391
QY	379	FIKETPGGVKEDMRIVKEITFGPVVTVTFKKSADSEVINMANSXYGLAAGTHTSINIFA	438
Db	392	FIKPTVFSEVDNNRIAKEELPGVPQILAKFSIEEVIKPSNSTDYGILTAAPVTNKILKA	451
QY	439	LKVAECRVNAGTIVGINTNDPHHAVPFGFNASGIGPMFSVLDQNTLQVKVAFKLI	494
Db	452	LKLSALESGVTINCYNALLYAAPPPTFKMSGNCPKIEYALAEAYTEVKVTFIKL	507

RESULT 9

RESULT 9
AA017364
ID AA017364 standard; protein; 512 AA.

[illegible]

XX	Sequence	512 AA
SO		

Query Match	47.8%	Score 1218.5,	DB 23,	Length 512;
Rest Total Similarity	50.4%	Pred No. 14E-94;		
Matches 240;	Conservative	84,	Mismatches 149;	Indels 3, Gaps 3

QY	22	TGLFINNNKVFVSKQNKTEFVILNPSTEEHETICHYGRRELDVEEAVQADRAFSNGS-WNGI	80
DB	32	TKLFINNNWHSSKSKKPKATNPSTPTPTPTFVPRFGDQDVKAVEAAVAFQPSWPR	91
QY	81	DPIDSKALYELARLIEFKFIVIASIETIGSKA-ISSPFDVFLVINYIKSSAGFAIKI	139
DB	92	DAI SGRII I HOLADI VFSDPATIIAAI ETMTDCKPFI HAFEDIDPGCTPTIYFAGWADKI	151

[illegible]

Qy 200 SALVSVKIPQAGIPPGVINIVSFGCKIIVVEALTNHPKIKKVAFTGSTATGPHIVQSA - 258
 :
Tb 212 TAYLIGSLKEAFPPGVYVINVPFGPTVTCAAISSHPDINKIAFGSTEVGFLVPEAASP 271

QY 259 ACLEKVVTELGGRSPNIVFADAEKKAVQNIIILGIYNSGHCVCCLSRVYVFESIVDKFI 318
 :
 Db 272 SNLKRVTLELGGKNPCIVCADADIDIAVFAHGVFFNGGCCCTAASFVFEQQVSEFV 331

```

QY      319 EEFKAASESIKVGDPFDESTFGQTSMLUNKILIKYVDLKNPGAFILFTGGERJASKG Y 378
          :: ||||| | : ||| : : |||| : |||| : ||
Db      332 RSVEYAKRRVPDPPDWKTQGGPQIDQKQFDKIILEIESCKFKFAKLFQGGSAMEFWGL 361

```

Qy 379 FIKPTVFGDVKEDMRIVKKEEIFGPPVVTVTKFKSADEVINMANDSEYGLAAGHTSNTINTA 438
 ||||| . ||| ||||||| |||| |||| |||| |||| |||| ||||
Dd 392 FIKPTVFSEVDNMRIAKEEIFGPVPILKFKEEVKANSTDYGLTAAVFTKNLDKA 451

RESULT 10
ABC06577

ABG06577
ID ABG06577 standard; Protein: 529 AA.

[illegible]


```

Lb 398 ATLQAGGDKLGGKGYIQPTVFSQVKIDMLIATDFIFGPVQITLKFDLDEVIARANSR 457
QY 424 YGLAAGIHTSINTALKAVDRVAGTWTINTYNDHFHAFVPPGFGFNASGLGRMSVDALON 483
Db 458 YGLAAGVFTONLDTAHLMDRALRVGTWVINCDFVLDASTPFGGKMGSGIGREKGIYSLNN 517
QY 484 YIQVKAVRAKL 494
Db 518 YIQVKAVTSL 528

RESULT 15
AAR63672
ID AAR63672 standard; Protein: 521 AA.
XX
AC AAPA3672;
XX
DT 08-MAY-1995 (first entry)
XX
DE Aldehyde dehydrogenase Aldh-1.
XX
KW Retro virus; vector, aldehyde-dehydrogenase;
KW glutamylcysteine-synthetase; hematopoietic cell, cyclophosphamide;
KW chemotherapy, transgenic animal, gene therapy, cancer therapy;
KW selectable marker.
XX
OS Homo sapiens.
XX
PN W09423015-A.
XX
PD 13-OCT-1994.
XX
PF 01-APR-1994; 94WO-US03624.
XX
PR 01-APR-1993; 93US-0041722.
XX
PA (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI Dalla-favera R, Gianni AM;
XX
DR WPI: 1994-333177/41.
DR N-PSDB; AAQ72450.
XX
PT Retroviral vectors encoding human cytosolic aldehyde
PT dehydrogenase or glutamyl cysteine synthetase - used to transform
PT a subject's hematopoietic cells to reduce the toxic effects of
PT cyclo phosphamide chemotherapy
XX
PS Disclosure; Fig.4: 92pp; English.
XX
CC A novel retro virus vector encodes human cytosolic aldehyde-
CC dehydrogenase and/or human glutamylcysteine-synthetase (AAR63673).
CC Hematopoietic cells transfected by the vector are resistant to
CC cyclophosphamide, providing a means of gene therapy that allows
CC higher doses of toxic drugs to be used in cancer chemotherapy.
CC The human genes may also be used as selectable markers for
CC mammalian cell transfection and for transgenic animal breeding.
XX
SQ Sequence 521 AA;

Query Match 45 28; Score 1153 5; DB 15; Length 521;
Best Local Similarity 48.5%; Pred. No. 4.7e-89;
Matches 234; Conservative 88; Mismatches 153; Indels 7; Gaps 5;

QY 20 QPTGLFINNKFPVKQNKTEVINPSTEEIICHIEGKEDLVBEAVQAAADRAFSNGS-WN 78
Db 35 QYTKIFINNEHDSVSGKKFPVFNPAEEELCOVEGDKEDVDKAVKAAQAFQIGSPWR 94
QY 79 GIDPIDRGKALYRIAELEQTK--DVIASIETLDNGKAISSS--PGVDVLVINLYKSSAGF 135
Db 95 TMDASEPGFLLYKLADLIERDLIATMESMESMNGKLYSNAYLNDLGGCIKTLRYCAGW 154
QY 136 AUKL--DGRMLIDIGKTHFSYIKRQPLGVGQQIIPWNPFLLMWAKIAPALVTGNIVLKT 193

```

```

Db 155 ADKLGGASPTIPILGNEFFIYTHPEKIGVGGQILPWNFPLVMLIKIGPAISGNIWVVKP 214
QY 194 AESTPIASIVVSYKTPGATGTPPQVINTIVSGFIVVEATINHPKIFKVAFTGSTATGRHI 253
Db 215 EEQTPLTALHVASLIEKAGFPFGVNVILPFGYPTAGAAISSHMDIDKVAFTGTFV3KLI 274
QY 254 YQSAA-AGLKKVILELGGKSNIVFAUAEELKRAVQNIILGIVYNSGEVCCAGSRVIVPES 312
Db 275 KEAAGKSNLRVLTLEUGGKSPCIVLADADLDNRAVEFAHGVFYHQGCCIAASRIFVERS 334
QY 313 IYDKFIEEPKAASESIKVGDPDFDESTFOGAQTQSOMOLNKLKYVDIGKNEGATLITGSHR 372
Db 335 IYDEFVPPSVERAKKYTLGNPLTPGVTOGPQIDKEQYDKILDLESKKKGALECGGGP 394
QY 373 LGSKGYEIKPTVFGDVKEDMPIVKEELFGPVVTVTFKSADEVINNMANDEYGLAACIHT 432
Db 395 WGNKGYEQPTVFSNVITDEMRIAKBEEIFGVQOIMKFKSLDDVDIKRANNTFYGLSAGVFT 454
QY 433 SNINTALKVALHPNAGIIVWINTYNCPIHIAVFPFGFNASGIGFPMSVIATQNYLQVKAVKA 492
Db 455 KDILKALITSSALQAGIVWVNYGVVSAAPFFGFKMSGNGRFLGEGFHEYTEVKTIVT 514
QY 493 KL 494
Db 515 KI 516

```

Search completed: June 24, 2003, 10:13:40
Job time : 41.8822 secs

1. The first part of the document discusses the importance of maintaining accurate records of all transactions and the role of the accounting system in providing reliable financial information. It emphasizes the need for transparency and accountability in financial reporting.

2. The second part of the document outlines the various methods used to collect and analyze financial data, including the use of statistical techniques and the application of mathematical models. It highlights the importance of using appropriate methods to ensure the accuracy and reliability of the results.

3. The third part of the document discusses the challenges faced by organizations in managing their financial resources and the role of the accounting system in addressing these challenges. It emphasizes the need for effective financial management and the importance of using the accounting system to monitor and control financial performance.

4. The fourth part of the document discusses the role of the accounting system in providing financial information to management and the importance of using this information to make informed decisions. It emphasizes the need for accurate and timely financial information and the role of the accounting system in providing this information.

5. The fifth part of the document discusses the role of the accounting system in providing financial information to external stakeholders and the importance of using this information to build trust and confidence. It emphasizes the need for transparency and accountability in financial reporting and the role of the accounting system in providing this information.

6. The sixth part of the document discusses the role of the accounting system in providing financial information to the public and the importance of using this information to make informed decisions. It emphasizes the need for accurate and timely financial information and the role of the accounting system in providing this information.

7. The seventh part of the document discusses the role of the accounting system in providing financial information to the government and the importance of using this information to make informed decisions. It emphasizes the need for accurate and timely financial information and the role of the accounting system in providing this information.

8. The eighth part of the document discusses the role of the accounting system in providing financial information to the media and the importance of using this information to make informed decisions. It emphasizes the need for accurate and timely financial information and the role of the accounting system in providing this information.

9. The ninth part of the document discusses the role of the accounting system in providing financial information to the public and the importance of using this information to make informed decisions. It emphasizes the need for accurate and timely financial information and the role of the accounting system in providing this information.

10. The tenth part of the document discusses the role of the accounting system in providing financial information to the public and the importance of using this information to make informed decisions. It emphasizes the need for accurate and timely financial information and the role of the accounting system in providing this information.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:15:55 ; Search time 13.70x5 seconds
(without alignments)
1062.435 Million cell updates/sec

Title: US-09-830-751-2
Perfect score: 2550
Sequence: 1 MSH1.PMTVPKIPKLPGLPEQ MSVDALQNYLQKAVPAKIL 495

Scoring table: BLASTUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA:
1. /cqn2_6/ptodata/1/iaa/5A.COMR.pep.*
2. /cqn2_6/ptodata/1/iaa/5B.COMR.pep.*
3. /cqn2_6/ptodata/1/iaa/6A.COMR.pep.*
4. /cqn2_6/ptodata/1/iaa/6B.COMR.pep.*
5. /cqn2_6/ptodata/1/iaa/ptus.COMR.pep.*
6. /cqn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	DB ID	Description
1	1158.5	45.4	521	4	US-09-221-294-2	Sequence 2, Appli
2	982.5	38.5	510	4	US-09-134-001C-4541	Sequence 4541, Ap
3	949.5	37.2	506	4	US-09-134-001C-4383	Sequence 4383, Ap
4	921.5	36.1	485	4	US-09-651-941-9	Sequence 9, Appli
5	921.5	36.1	485	4	US-09-955-547-9	Sequence 9, Appli
6	921.5	36.1	508	4	US-09-655-270A-4	Sequence 2, Appli
7	888	34.8	497	1	US-08-513-841-2	Sequence 2, Appli
8	888	34.8	497	2	US-08-696-834-2	Sequence 2, Appli
9	888	34.8	497	2	US-08-942-674-2	Sequence 2, Appli
10	888	34.8	497	4	US-08-118-317-2	Sequence 2, Appli
11	775.5	30.4	488	4	US-09-144-001C-434F	Sequence 434F, Ap
12	763.5	29.9	493	4	US-09-134-001C-4388	Sequence 4388, Ap
13	685.5	26.9	518	4	US-09-134-001C-4451	Sequence 4451, Ap
14	684	26.8	487	4	US-09-351-224F-5	Sequence 5, Appli
15	657.5	25.8	482	4	US-09-155-183-4	Sequence 4, Appli
16	365	14.3	464	4	US-09-134-001C-4701	Sequence 4701, Ap
17	338.5	13.3	133	1	US-08-346-611-2	Sequence 2, Appli
18	338.5	13.3	133	2	US-08-794-494-2	Sequence 2, Appli
19	194.5	7.6	551	4	US-08-952-661-2	Sequence 2, Appli
20	152	6.0	711	4	US-09-114-001C-544-0	Sequence 544-0, Ap
21	124.5	4.9	1002	1	US-09-268-347-24	Sequence 24, Appli
22	115.5	4.5	510	1	US-08-489-733-1	Sequence 1, Appli
23	115.5	4.5	510	2	US-08-993-581B-1	Sequence 2, Appli
24	115.5	4.5	510	4	US-09-134-078-22	Sequence 22, Appli
25	108.5	4.3	255	4	US-08-858-207A-4F5	Sequence 465, App
26	104	4.1	495	4	US-09-134-001C-3128	Sequence 3128, Ap
27	164	4.1	1433	4	US-09-376-330-16	Sequence 16, Appli

28 102 4.0 72 4 US-09-655-270A-33 Sequence 33, Appli
29 102 4.0 509 4 US-09-134-078-18 Sequence 18, Appli
30 100.5 4.4 555 2 US-09-477-07A-6 Sequence 6, Appli
31 99.5 3.9 2314 4 US-09-268-347-49 Sequence 49, Appli
32 98 3.8 687 1 US-08-164-839-31 Sequence 31, Appli
33 98 3.8 587 1 US-08-164-839-31 Sequence 33, Appli
34 98 3.8 687 1 US-08-583-799-31 Sequence 31, Appli
35 98 3.8 527 1 US-08-583-799-33 Sequence 33, Appli
36 98 3.8 588 1 US-08-164-839-70 Sequence 70, Appli
37 98 3.8 588 1 US-08-164-839-72 Sequence 72, Appli
38 98 3.8 688 1 US-08-583-799-70 Sequence 70, Appli
39 98 3.8 588 1 US-08-583-799-72 Sequence 72, Appli
40 98 3.8 2404 4 US-09-134-001C-3464 Sequence 3464, Ap
41 97.5 3.8 464 4 US-08-984-618-13 Sequence 13, Appli
42 97.5 3.8 495 4 US-09-217-490-2 Sequence 2, Appli
43 97.5 3.8 662 1 US-08-224-657-94 Sequence 94, Appli
44 97.5 3.8 662 4 US-09-354-138-94 Sequence 94, Appli
45 97.5 3.8 1053 4 US-09-724-510-2 Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-221-294-2
; Sequence 2, Application US/09221294
; Patent No. 6268138
; GENERAL INFORMATION:
; APPLICANT: Riccardo Dalla-Favera and
; APPLICANT: Alessandro Massimo Gianni
; TITLE OF INVENTION: A Retroviral Vector Capable of Transducing the
; TITLE OF INVENTION: A Retroviral Vector Capable of Transducing the
; TITLE OF INVENTION: A Retroviral Vector Capable of Transducing the
; TITLE OF INVENTION: A Retroviral Vector Capable of Transducing the
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER PREPARED FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM 330 466 DX2
; OPERATING SYSTEM: PC-PMOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221,294
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DRAWING NUMBER: 4,940-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-291-0525
; INFORMATION FOR SEQUENCE CHARACTERISTICS:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 521 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-221-294-2

Query Match 45.4%, Score 1158.5, DB 4, Length 521;
Best Local Similarity 48.8%, Pred. No. 8,23-100,
Matches 235, Conservative 88, Mismatches 152; Indels 7; Gaps 5;
QY 20 QPTGLFINNKVPKQKTEVINPSTEEETIHYEEDVEAVQAAAPAFSGS-WN 78
DE 35 QVTKTFINNEHDSVSKKFFVFNFATEERELQVEEGCKEDV-KAVKAAQAFQETGSPW 94
QY 79 GTPDTPKALYPLAEILLQK--IVIASIEILLNCKAISSE-PCIVDIVINYKSSAGF 135


```

; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4246
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-4246

Query Match      30 4%, Score 775.5; DB 4; Length 488,
Best Local Similarity 36.6%; Pred. No. 56-64;
Matches 174; Conservative 97; Mismatches 188; Indels 17; Gaps 7;

QY 22 TLEFNKVFVSKNKTFFVNFSTEEIECHYGFPGVGEAAVQAALRAFSNWSN:110 H1
DB 18 TKYINGEWDVSASGETIDVINPATIEVMGKIAGNEEDVNKAADAPKVFYLFPHSSVE 77
QY 82 PIDRCKALYRLAELEQDKVIASTFTLEWCKATSSSPGIVIVINY...LKSSACFAUK 138
DB 78 --ERRELLDKIVKEYQNKNDLIPATTEIGAGVLSN-----VHYQMSLNHFTAARQA 130
QY 139 IDGRMDTGRTHFSYKPEPLVAGVGGIIPWNEPILMAMWAKIAPALVTCNTVVLKTAESTP 198
DB 131 LDSFGFEPQKGLU-DLVKKEALGVAGLVTPWNEPINOISLKAAFAAGSHVVLKPSSETP 189
QY 199 LSALYVSKYIPQAGLIPGVINVSFGFKIVVEATINHPKIKKVAFTGSTATGRHIYQSA 258
DB 190 FAAILIARIFDKVGPVGVNLVNGDGVGNPLSEHPKVRMMSFTGSGTSGKIMEKAA 249
QY 259 AGLKKVTELEGGKSPNIVFADELKKAQVNTILGIVYNSGECVGCAGSRVYVESIYDKFI 318
DB 250 KDFKVSLELGGKSPYIVLDVVDVEEAANATIKKVVNNTQVCTAGTRVLIPESIKEDYL 309
QY 319 EEFKAASIESIKGDFPESTFGAGTTSQMUNIKILKYVIGKNEGATILIGG---ERLG 474
DB 310 TAVKRAFSKVRVGPQPREKGTQVGPILSKKQFDQVQYILKGIINBTAELPVGSPKPKENL- 468
QY 375 SKGYEIKPTVFGDKEDMRIVKKEIFGVVTVTKFKSADRVINMANSYGLAAGIHTSN 434
DB 366 URGYFAPPTIFINLNMHTAQLPEIFPVMSVITYNNLDAIEIANTRYGLAGYVIGKD 428
QY 435 INTALKVADRVNAGTVINITYNDFHVAVPPGFGFNASGLGREGMSVIALNYLAVKAV 490
DB 429 KDTLPHVARSTEAETIEINEAGR-KPELPFGGKESGLGPHFWYGIIEFLEVKS1 483

RESULT 12
US-09-134-001C-4388
; Sequence 4388, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4388
; LENGTH: 493

```

```

; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-4388

Query Match      29.9%, Score 763.5; DB 4; Length 493;
Best Local Similarity 34.5%; Pred. No. 6,80-63;
Matches 162; Conservative 109; Mismatches 181; Indels 17; Gaps 6;

QY 24 LFINNKVFVSKNKTFFVNFSTEEIECHYGFPGVGEAAVQAALRAFSNWSN:110 H1
DB 24 LFINNFIQSQSKETMDVINPATIEAFDITITLAEVEVNDIAEKSOOA--QLFWERKVP 81
QY 84 DPGKALYPLAELEQDKVIASTFTLEWCKATSSSPGIVIVINY...LKSSACFAUK 141
DB 82 TRACHVKLLIPLEKKNKDELAOLYVKGKILAGAGELIDKISFIDYMTLSMSUK--G 139
QY 142 PMIDRPFTHFS-YTKPQPDVAGVGGIIPWNEPILMAMWAKIAPALVTCNTVVLKTAESTP 200
DB 140 PVLQNSIANETIQIINKPIGVTAIVPWNAPILVLMKVPITPAIVTGVSVVVKPSEPTTIL 199
QY 201 ALYVSKYIPQAGLIPGVINVSFGFKIVVEATINHPKIKKVAFTGSTATGRHIYQSA 250
DB 200 TIFLAELFPASTIPALVGVIGVGGTFTVTCGLASHKELFETISLDSMPAFKSVYPAVAT 259
QY 261 LKVTITELGKSPNIVFADELKKAQVNTILGIVYNSGECVGCAGSRVYVESIYDFEIEE 320
DB 250 VKKNVLELQGNAPVIVISNALLKFAVNYIVTAPINAGVCTCTGEPFVIEVWHDQFINP 419
QY 341 PKAASESIKVGDFPESTFGAGTTSQMUNIKILKYVIGKNEGATILIGG---ERLG 474
DB 340 VTSKMSLSLVGDPEDENTLVCATILNQGLDSIHEKVQDAIFKNCATILMTGCHQIKPHPEY 479
QY 381 KPTVGVGVKEDMRIVKKEIFGVVTVTKFKSADRVINMANSYGLAAGIHTSN:110 H1
DB 380 APTVLDNRKDYNVFKDEIFGPVLATITTYRDFEQVIDANDTNGAGUSSYIFSLENLEVM 439
QY 441 VADRVNAGTVIN-----TYNDFHVAVPPGFGFNASGLGREGMSVIALNYLAVKAV 484
DB 440 ATERLKPGEVYANCEAEVWNGYH-----AGWRESGLGGADGTHGPEY 483

RESULT 13
US-09-134-001C-4451
; Sequence 4451, Application US/09134001C
; Patent No. 6480470
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4451
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-4451

Query Match      26.9%, Score 585.5; DB 4; Length 518;
Best Local Similarity 33.1%; Pred. No. 1,5e-55;
Matches 173; Conservative 90; Mismatches 200; Indels 51; Gaps 14;

QY 6 MTPVPIKPLNGLEYEQPTG-----LFINNKVFVSKNKTFFVNF:44
DB 5 MWVPEKNEPGLDFSVQTNVERFNEBELPKVKAQLQJQJVLVINGEKL--TKIDTFNSVIA 62
QY 45 STEEIECHYGFPGVGEAAVQAALRAFSNWSN:110 H1
DB 45 STEEIECHYGFPGVGEAAVQAALRAFSNWSN:110 H1

```

```

Db      63 NTSQIAKVSQATODDIFKAFESANHAYG--SHKWSHKDRAELLRLVAAIIIPRKEELS 120
QY      105 SIETLONKALISSRGD-----VDLVINYLKSSAGFADKIDGR-MIDTGRTHFSYTKROPL 159
Db      121 AIMVYEAGKPWDEAVGDAAGIDFIEYARSMELA---DGRPVLDREGEHNRHYEK-PI 176
QY      160 GVCQGIIPWNPPIILMAWKIAPALVTGNVWLKTAESIFLSALYVSKYIPQAGIIPGVIN 219
Db      177 GTGVTIPWNPPIAIMAGTTLAPVAGNVTLLKPAEDTVLTAYKMLEIEEAGLPGGVN 236
QY      220 IVSFGKIVVEAITNHPKIKKVAFTGSTATGRHIYQSA--GLKKVTLLELGGKSP 273
Db      237 FVPDGPKEIGDYLDVHDHKTDFVTFTGSRATGKIYERSAVVQGGOFKRVIAEMGGKDA 296
QY      274 NIVEFAELKAVONIIILGIYNSGEVCCAGSRVVEESIDYKFIIEFKAASESIKVQDP 333
Db      297 IVDNNVDTLLAAEIVTSAGFSQKQCSACSRAIVHQVHDEILEKAIQLTKLILGNI 356
QY      334 FDESTFOCAOTSQMLNKILKYVDIGKNFEGATLITGGERLGSKGIFYIKPTVFGVKEDMR 393
Db      357 -EENTFMGPVINOQFDKIKNYIEIGKKEG-KLETGGTDDSTGYFIEPTIFSGLOSADR 414
QY      394 IVKEIEGPVVTYTKPSADEVINMANDSEYGLAAGIHTSNINTALKVADRWNAGTWIN 453
Db      415 IMQEEIFGPVVGFTKVKDFDEALEVANDTDYGTGAVITNHRHFWIKAVNEFDVGNLYIN 474
QY      454 ----TYNDFHVAVPFGGFNASGL-GREMSVDALONYLOVKAV 490
Db      475 RGCTAAVVGYH---PFGGFKMSGTDAKTGSPDYLLNFIEQVV 514

```

RESULT 14

```

US-09-351-224E-5
; Sequence 5, Application US/09351224E
; Patent No. 6388171
; GENERAL INFORMATION:
; APPLICANT: Duvick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald P.
; TITLE OF INVENTION: Compositions and Methods for Fumonisin
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/351,224E
; CURRENT FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSPQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-351-224E-5

```

```

Query Match      26.8%, Score 684; DB 4, Length 487;
Best Local Similarity 33.4%, Pred No. 18, 55;
Matches 164; Conservative 98; Mismatches 194; Indels 36; Gaps 12;

QY      19 EOPTGLFINNKPVPSSKONKTFEINPSTEEIEC-HIYEGREDDVBEAVOAAADRAFSGWS 77
Db      7 EYKSELFINNEFFVSSGSERLITLNPWFSTVAIDVHVANAADVSAASVQAKKQPV 66
QY      78 NGIDPIDRKALYRIAEILPQDKDVIASTETLONKALISS-SRGVDLVINYLKSSAGFA 136
Db      67 KKEFTAGRAACMLKPADLAEKNAEKLALESLEPTGRPVSMITHFDIPNWSVFVYAGWA 126
QY      137 DKIDGRMI--DITGRTHFSYTKRQPLGVCQGIILPWNPEILMAWKIAPALVTGNVWLKTA 194
Db      127 DKTAGTTFEDNCKPNWRY---EPMGVCAGIASMNATELYGVCKIAPALAAAGCSFIFKAS 183
QY      195 ESTPLSALYSKYIPOAGIIPGVINIVSGFKIVVVEAITNHPKIKKVAFTGSTATGRHIY 254
Db      184 EKSPGLVGIAPLAFAPAGPPQWQFIIG-APVTCEALASHMDIAKISFTSRVSGGQRAVK 242

```

```

QY      255 QSA-AAGLKKVTLLELGGKSPINIVFAIAELKKAQVONIIILGIYNSGEVYCAISNR 410
Db      243 QATIKSNMKVTLLELGGK-PTIVFNEAPLER--QSSESARQTSFPGQIWDV 290
QY      311 EST---YDKFIEFKAASESIKVG-DPDESTFOCAOTSQMLNKILKYVDIG 359
Db      291 ESTLIVQWENLAKKFEVPHPSFGQGLWMLQNDPFGKPIIDVPIKQYDRVLTGNLGG 350
QY      360 KNEVATLILIGGERLGSKGIFYIKPTVPGVDEKIMRIVKEEIPGVVTVITFKPSADEVDMA 419
Db      351 K-DIAGLILGVSKGKKGATIEDPIIVNKKPSKIFWEEFPGWLSIKTERFEFAIEFA 409
QY      420 NDSEYGLAAGIHTSNINTALKVADRWNAGTWINYNDHVAHPROGDNASGLGDEMSVD 479
Db      410 NQITYGLASVITYTKSLNPKIIVSSALFTGVSINFPFPIFPIFGMKQSSGRELAGE 469
QY      480 ALONYLOVKAV 490
Db      470 GIKAYLEPKTI 480

```

RESULT 15

```

US-09-155-183-4
; Sequence 4, Application US/09155183
; Patent No. 6324011
; GENERAL INFORMATION:
; APPLICANT: Narbad, Arjun
; APPLICANT: Rhodes, Michael J.C.
; APPLICANT: Gasson, Michael J.
; APPLICANT: Walton, Nicholas J.
; TITLE OF INVENTION: PRODUCTION OF VANILLIN
; FILE REFERENCE: 20747/100
; CURRENT APPLICATION NUMBER: US/09/155,183
; CURRENT FILING DATE: 1999-05-03
; EARLIER APPLICATION NUMBER: PCT/GB97/00809
; EARLIER FILING DATE: 1997-04-24
; EARLIER APPLICATION NUMBER: GB96/06187
; EARLIER FILING DATE: 1996-03-24
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Pseudomonas fluorescens
US-09-155-183-4

```

```

Query Match      25.8%, Score 657.5; DB 4; Length 482;
Best Local Similarity 32.9%, Pred. No. 5, 40, 53;
Matches 156; Conservative 98; Mismatches 205; Indels 15; Gaps 8;

QY      24 LFINNKEVPSKONKTFEIVINPSTEEETPHIYETREDFWEFAVAAARAFSSNNGHDP 83
Db      6 LILGGGQCPARQGRTERENPVTEGLVSKVAATLIEDAIAVAAAGQAF 63
QY      84 IPKALYPIAEILIEDVFIASIEDLNTKALISSSG-EDLVINYLSACFAIKDIP 142
Db      64 ERESILLKAAEQLOARSGEF--TEAGETGAMANNYGFNVRLAANMLREAAASMTIYVNGE 121
QY      143 MDTIG-PTHEFSYTKRQPLGVCQGIILPWNPEILMAWKIAPALVTGNVWLKTAESTPLSA 201
Db      122 VIISVPGSFAMLRQPCGVVGLIAWNAVPVLTATRAIMPLATUNIVVKASELSAVH 181
QY      202 LVYSYFPAAGLTPSVTRFVSGI--GFVVLATINHPKIFEVATFESTATPDEHYGAA 258
Db      182 RLIGGVLOIAGLGDGVVNVISNAPALAAQIVRLIANIIVRRVNTSTSTHVRVIGELSA 241
QY      259 AGLKKVTLLELGGKSPINIVFAIAELKKAQVONIIILGIYNSGEVYCAISNRVVEESLYEKF 418
Db      242 RHUKFALLELGGKAPILLIDRDLAEAAVAAAFAGVFNQGIIPKSTERLIIVAKVAFV 401
QY      319 EEFKAASESIKVGDFPDESTFEGAGTISQMLNKILKYVCKNKGALILIGGERLGSKG 478

```

Db 302 AQLAAKVETLRAGDPADESVLSIVLASACTRIKALIDUAVAKGALVIGGQLEGS--- 358
QY 379 FIKPTVFGDVKEDMRIVKEEIFGPVTVTKREKSADDEVINMANUSEYGLAAGIHTSNINTA 438
::||: | ||: || |||| | : : : ||||: ||: | : |
Db 359 IIQPTLLDVTASMLYFEESFGPVAVVLEFGFEFALIQIANDSFFGLSAAIFSPDTGPA 418
QY 439 LKXADRVNAGTWIN--TYNDFHHAVPFGGFNASGLGREMSVDALQNYLOVKAV 490
| : | | : | || | : | : |||| | || : ||||: | : : |
Db 419 LALAORVESGICHINGPTVHD-EAOMPFGGVKSSGGYSGFGGKASIEHFTOLRWV 471

Search completed: June 24, 2003, 10:31:38
Job time : 15.7085 secs



Genome version 5.1 K
Copyright (c) 1993 - 2003 Cambridge

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:17:21 : Search time 23.9275 Seconds
(without alignments)
2238.528 Million cell updates/sec

Title: US-09-830-751-2

Sequence: 1 MSHLPMTVPIKUPNILEYEL MSVIALQNYLVKAVPAKLD 495

Scoring table: MS-09-830-751-2

Gapop 10 0 0, Gapext 0.5

Searched: 41779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 41779

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 100%

Listing first 45 summaries

Database: Published_Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUB pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUB pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_PUB pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUB pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUB pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUB pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US06_PUB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1399	54.9	496	US-09-847-208-59	Sequence 59, Appl
2	1362.5	53.4	495	US-09-847-208-11	Sequence 11, Appl
3	1218.5	47.8	512	US-09-961-403-12	Sequence 12, Appl
4	1218.5	47.8	512	US-10-268-518-2	Sequence 2, Appl
5	1218.5	47.8	512	US-10-305-823-14	Sequence 14, Appl
6	1201	47.1	538	US-09-344-882-20	Sequence 20, Appl
7	1201	47.1	538	US-10-293-865-20	Sequence 20, Appl
8	1169.5	45.9	534	US-09-344-882-24	Sequence 24, Appl
9	1169.5	45.9	534	US-10-293-865-24	Sequence 24, Appl
10	1146.5	45.0	501	US-09-444-882-22	Sequence 22, Appl
11	1146.5	45.0	501	US-10-293-865-22	Sequence 22, Appl
12	1067	41.8	492	US-10-268-518-4	Sequence 4, Appl
13	1056.5	41.4	493	US-10-175-696-21	Sequence 21, Appl
14	1056.5	41.4	493	US-09-823-901-9	Sequence 9, Appl
15	1025	40.2	496	US-09-815-242-10550	Sequence 10550, A
16	1015	39.8	508	US-08-781-986A-5241	Sequence 5241, Ap
17	984	38.6	500	US-10-166-087-4	Sequence 4, Appl
18	970.5	38.1	496	US-09-815-242-5644	Sequence 5644, Ap
19	970.5	38.1	496	US-09-815-242-12657	Sequence 12657, A

20	948	36.8	490	US-09-815-242-1262	Sequence 1262, A
21	929	36.0	490	US-09-815-242-10657	Sequence 10657, A
22	899	34.1	505	US-09-738-626-6572	Sequence 6572, Ap
23	878	34.5	412	US-09-928-302-736	Sequence 736, App
24	878	34.4	518	US-09-919-039-143	Sequence 143, App
25	845.5	32.8	487	US-10-175-696-17	Sequence 17, Appl
26	845.5	32.8	487	US-09-823-901-5	Sequence 5, Appl
27	845.5	32.8	487	US-09-925-350-148	Sequence 148, Ap
28	825.5	32.4	481	US-09-815-242-13829	Sequence 13829, A
29	781.5	30.6	475	US-09-815-242-12475	Sequence 12475, A
30	761.5	30.0	482	US-09-815-242-18264	Sequence 18264, A
31	761.5	30.0	482	US-09-748-626-4108	Sequence 4108, Ap
32	761.5	30.0	482	US-09-815-242-6058	Sequence 6058, Ap
33	754.5	29.6	482	US-09-815-242-14047	Sequence 14047, A
34	735	28.8	490	US-09-748-626-3558	Sequence 3558, Ap
35	586	25.3	481	US-09-748-626-6440	Sequence 6440, Ap
36	584	25.8	487	US-09-882-994-5	Sequence 5, Appl
37	562	26.0	491	US-09-997-664-3	Sequence 3, Appl
38	557.5	25.8	482	US-09-733-383-4	Sequence 4, Appl
39	551.5	25.5	498	US-10-272-419-12	Sequence 12, Appl
40	537	25.0	481	US-08-976-0630-28	Sequence 28, Appl
41	622	24.4	521	US-09-738-626-6482	Sequence 6482, Ap
42	591.5	23.2	439	US-09-815-242-13316	Sequence 13316, A
43	584	23.1	483	US-09-748-626-4047	Sequence 4047, Ap
44	579	22.7	584	US-09-748-626-3680	Sequence 3680, Ap
45	472	18.5	852	US-09-838-541-2	Sequence 2, Appl

ALIGNMENTS

FFSULT 1
US-09-847-208-59 Application US/09/847208
; Sequence 59, Application US/09/847208
; Publication No. US20030082150A1
; GENERAL INFORMATION:
; APPLICANT: Saxon, Andrew
; APPLICANT: Zhang, Ke
; APPLICANT: Zhu, Daoheng
; TITLE OF INVENTION: FUSION MOLECULES AND TREATMENT OF
; TITLE OF INVENTION: IgE-MEDIATED ALLERGIC DISEASES
; FILE REFERENCE: CC67-003A
; CURRENT APPLICATION NUMBER: US/09/847,208
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 177
; SOFTWARE: FastSPQ for Windows Version 4.0
; SEQ ID NO 59
; TYPE: PRT
; ORGANISM: Cladosporium herbarum
US-09-847-208-59

Query Match	54.9%	Score 1399	PR 9	Length 456
Best Local Similarity	53.8%	PR 9	4.4e-98	
Matches	263	Conservative	80	Mismatches 135
				Indels 2
				Gaps 2
27	7	IVLIFLNGSLYFTGLFINKKVFVKONKIFEVINFTSTEEICHIVYEDRWVAVQ	66	
Db	3	SVOLETHSGKYEQFTGLFINNEFKQCKFDVINSDESVITGVHEATKIKVDIAVA	62	
27	67	AAHAFNSNSWNSLPHTRCKALYHIAFLLECKKVAASIFLITLNTKATSSSPGLDLV	126	
28	63	AAEAF-EGSWRLTEFENCKLNNLANFLERNTLLAAVESIDNGKATSMFVTSACAS	121	
27	127	NYLKSSACFARKINGMIRGTFTFSTKPKQPGVGGQIIFNNEFLIMWVIAIALVIG	186	
28	122	GLLYFYSWAKITGKVLGTFTFTPNVYKFKPQVTPSQUSLPLIMWVIAIALVIG	181	
27	187	NTVILKATSPISALYVSKYIPKACIPPCVINVSFQKIVVEATINHKKKVAFQCS	246	
28	142	NTVILKATSPISALYVSKYIPKACIPPCVINVSFQKIVVEATINHKKKVAFQCS	241	
27	247	TATCPHYQSAAGLKKVTLFLGCRSPNIVFAFAHKKAVQNIILLCIYYSFVQVQNS	305	

```

Db      242 TVGRTILKAAASNLKVTLEUGKSPNIVFEDADIDNAISWVNFIFNHHGCCAGS 301
QY      306 RVVVEFSYDKFIEEKAASESIKVGDPDESTFQGAQTSOMGLAKILKYVDIGNEGAT 365
Db      302 RVVVEFSYDKFKVQKKEKAKVNVGUPPAADTFQSPQVSKVQFQKIMFYIQAGKIAGAT 361
QY      366 LITGGERISKGKGYFKPTVPGVKKEDMPVIVKEEIFGPVVVTVFKPSADEVINMANSSEYG 425
Db      362 VETGSRKSGKGYFIEPTIFSNVTEDMKIVKEEIFGPVCSIAKFKTKEDAIKLGNASTYG 421
QY      426 LAAAGHTSNINTALKVADRVNAGTVWINTYNDFHAAVDFEGFNASGLGKREMSVVALONYL 485
Db      422 LAAAVHTKNINTALEVSNALKAGTVWVNTYNTLHHOMPEGGYKESGIGRELGEDALANT 481
QY      486 QVKAARAKL 494
Db      482 QIKIVSIRL 489

RESULT 2
US-09-847-208-11
; Sequence 11, Application US/09847208
; Publication No. US20040082190A1
; GENERAL INFORMATION:
; APPLICANT: Saxen, Andrew
; APPLICANT: Zhang, Ke
; APPLICANT: Zhu, Daoheng
; TITLE OF INVENTION: FUSION MOLECULES AND TREATMENT OF
; FILE REFERENCE: IgE-MEDIATED ALLERGIC DISEASES
; CURRENT APPLICATION NUMBER: US/09/847,208
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 177
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 495
; TYPE: PRT
; ORGANISM: Alternaria alternata
US-09-847-208-11

```

```

Query Match      53.4%; Score 1362.5; DB 9; Length 495;
Best Local Similarity 54.3%; Pred. No. 2.6e-95;
Matches 266; Conservative 84; Mismatches 135; Indels 5; Gaps 5;

QY      7 TVPKLPNGLEYDPTGLFINNKVFESKONKTFEVINISTEEFICHYEGREDIVFEAVQ 66
Db      3 SVKLSPTQTFPQPTGLFINNEFKVAVGKTFDVLNISTEEVCSVOEATEKDVDAVA 62
QY      67 AADAFNSGNSWGIDPTPKKALYLAELIEGKIVIASIETILUNGKAISSRG-DVULV 125
Db      63 AAKAF-NPWAKEPENNKGKLIINKLADLFEKNADILAAVEALDNGKAFSNKAVDVPAA 121
QY      126 INYKSSAGFAUKIDGKMDTSTHESYTKRQPLGWVQJIPWNEPILMWANKIAPALVT 185
Db      122 AGCLPYVGGMAIKFKKVVITAPDSNYIY-RKSLVFAVRSSMELPILMWSKICPAIAT 180
QY      186 GNTVLKLTARSTPLSALYVSKYIPQAGIPPVINIVSGFKIVFAITNHFKIKKVAFTG 245
Db      181 GNTVLKLTARSTPLSALYVSKYIPQAGIPPVINIVSGFKIVFAITNHFKIKKVAFTG 243
QY      246 STATGRIHYOSAA-AGLKVVTELGKSNIVFAIAELIKKAVQNTILGITYYNSGRCVAG 304
Db      241 STVVGQIMKSAAGSNLKVVTELGKSNIVFAIAELIKKAVQNTILGITYYNSGRCVAG 303
QY      305 SRVYVESIVDKFIFBEKAASESIKVGDPDESTFQGAQTSOMOLINKILKYVDIGKNEGA 364
Db      301 SRIYVQFEIYDKFIQRFKRAAONAVGDPF-AATIQGQVQVQJQFDIRIMYTFESKSGA 359
QY      365 TLITGGERISKGKGYFKIPTVFGDVKKEDMRIVKEEIFGPVVTVTKFKSADEVINMANSSEY 424
Db      360 TIETGGRKSGKGYFIEPTIFSNVTEDMKIQOFEIFGVPVCTISKFKTKAIVIKIGNNTY 419

```

```

QY      425 GLAAGHTSNINTALKVADRVNAGTVWINTYNDFHAAVDFEGFNASGLGKREMSVVALONY 484
Db      420 GLSAAVHTSNLITAEVANALRAGTVVNSYNTLHWOLFPGGYKESGIGRELGEDALANTY 479
QY      485 IAVKAVPAKL 494
Db      480 IOTKTVSIRL 489

RESULT 3
US-09-961-403-12
; Sequence 12, Application US/09961403
; Publication No. US20030077589A1
; GENERAL INFORMATION:
; APPLICANT: HE-STUMPP, HOLGER
; APPLICANT: HAENDLER, BERNARD
; APPLICANT: KRAETZSCHMAR, JOERN
; APPLICANT: KREFT, BEKTHOLT
; APPLICANT: WINTERHAGER, ELKE
; APPLICANT: RECIDOR, PEDRO
; APPLICANT: SCOTTI, SIMONE
; TITLE OF INVENTION: METHOD FOR IN VITRO DIAGNOSIS OF ENDOMETRIOSIS
; FILE REFERENCE: SCH-1789
; CURRENT APPLICATION NUMBER: US/09/961,403
; CURRENT FILING DATE: 2001-09-25
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 12
; LENGTH: 512
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-961-403-12

```

```

Query Match      47.8%; Score 1218.5; DB 9; Length 512;
Best Local Similarity 50.4%; Pred. No. 2.2e-84;
Matches 240; Conservative 84; Mismatches 149; Indels 4; Gaps 4;

QY      22 TGLFINNKVFESKONKTFEVINISTEEFICHYEGREDIVFEAVAAURAFSNIS WNGI 80
Db      32 TRIFTNNEWHESSKSKKFKATCNSTREQICEVEEDKPPVDKAVIAAVAFQKRSWPR 91
QY      81 DPIDGKALYLAELIEGKIVIASIETILUNGKAISSRGIVIVLYNLSKSAEALPT 139
Db      92 DALSKGLLHQLADLVEROKALIALETMDIGKPELHAFPTDLERCTILAYFAWALPT 151
QY      140 DGRMILITGTHESYTKRQPLGWVQJIPWNEPILMWANKIAPALVTGNTVLKTAESTPL 199
Db      152 QCKTITIDUNVVTFRHEPTFGVCAITWNTTILMWKILALALASGRIMVIRIADQPL 211
QY      200 SAIYVSKYIPQAGIPPVINIVSGFKIVFAITNHFKIKKVAFTSTATGRIHYOSAA 258
Db      212 TALYLASTIKEN-FTDQGVNIVPQPTVGAALSSHPG-INPILATGSEV-KVFAA-P 271
QY      259 AGLKVVTELGKSNIVFAIAELIKKAVQNTILGITYYNSGRCVAGSVVYVESIYKPT 318
Db      272 SNLKVVTELGKSNIVFAIAELIKKAVQNTILGITYYNSGRCVAGSVVYVESIYKPT 341
QY      319 BEFKAASESIKVGDPDESTFQGAQTSOMOLINKILKYVDIGKNEGA 378
Db      332 PFSVYAKKPPVQVTEPKKELGQVQJQFDIRIMYTFESKSGA 391
QY      379 FKPTVFGDVKKEDMRIVKEEIFGPVVTVTKFKSADEVINMANSSEYGLAAGHTSNINTA 438
Db      392 FKPTVFESEIUNMKIATKFEIFGPVQFILKESFEVIRKANSIY-ILAAVE-ENLKA 451
QY      439 LKVAQVHNAGTVWINTYNDFHAAVDFEGFNASGLGKREMSVVALONYLQVKAARAKL 494
Db      452 LKLASLESFTVWINTYNAIYADAPFEGKMSGNGPELGAYALAFYTFVTVI 507

```

```

PSFMT 4
US-10-268-518-2
; Sequence 2, Application US/10268518

```

```

; Publication No. US2003010003A1
; GENERAL INFORMATION:
; APPLICANT: Hunter, John Joseph
; TITLE OF INVENTION: 5136, A HUMAN ALGEBYRE DEHYDROGENASE
; FILE REFERENCE: MPT01-234P1RM
; CURRENT APPLICATION NUMBER: US/10/268,518
; PRIOR FILING DATE: 2002-10-10
; PRIOR APPLICATION NUMBER: 60/429,899
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 512
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-268-518-2

Query Match          47.8%; Score 1218.5; DR 9; Length 512;
Best Local Similarity 50.4%; Pred. No. 2 de-84;
Matches 240; Conservative 84; Mismatches 149; Indels 3; Gaps 3,

QY 22 TGLFINNKVFVSKNKTFEVINSPREELCHIVGCPHUUVEEAVQAADRAFSNCS WNCI 80
DB 32 TKIFINNHESKSKKFAFNPSTREQICVVRGCHUKPIVRAKAAQVAFQGSWPKL 91
QY 81 DPIDCKALYPLAELEQCKIVIASIFETLNGKA-ISSSPGVULVINYLKSSAGADKI 139
DB 92 DALSGRLLHULALIVKRAIALALETMTGKPFTHAFFIDLEGCIPTLRYFAGWADKI 151
QY 140 DPMIDGTHFSTYKQPLGVCCGLIPWNPFLIMMAWKIAPALVIGNTVVLKTAESTPL 199
DB 152 QCKTPTDNNVCTFHEPFGVCAITPWNPFLIMVWKIAPALCGNTWVTKPAQTPL 211
QY 200 SALYVSKYIPGAGITPGVINIVSGKLVVEATINHKIKKVAFTSTAGRIHYOSAA- 258
DB 212 TALYLGSLIKAGFPGVNVIVGFGTVAATSSHPQINKIAFTGSTEVCKLVKKAASR 271
QY 259 AGLKVTLELGGKSNIVAFAPAEKLVKAVQNTILGTYVNSGEVTCAGSPVVEESIVDKFI 318
DB 272 SNLKRVTELOCKNFCVCAADLDLAVEAHQGVFFNCGCCCTAASPRVFEQVYSEFV 331
QY 379 FIKPTVFGVADKMKLVKEEFSPVIVIKPKSADEVINMANUSEYGLAAGIHTSNINTA 438
DB 392 FIKPTVSEVTONMPLAKEEFPGVQPLIKPKSTEEVIKKNSTIDVGLTAAVETKNLKA 451
QY 439 LKVADEVNAGTVWNTYNIPHHAVPPGSPNASISIGRMSVIALQNYLQVAKVAPKL 494
DB 452 LKLASALESGIWINCYNALLYAQAPFGGFKMSGNGRELDYALAEYTEVKTIVIKL 507

RESULT 5
US-10-205-823-14
; Sequence 14, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Willson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Womsey, Angela M.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND

```

```

; TITLE OF INVENTION: THERAPY OF POST-TATE CANCER
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; PRIOR FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/425,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 512
; TYPE: PPT
; ORGANISM: Homo sapiens
US-10-205-823-14

Query Match          47.8%; Score 1218.5; DB 9; Length 512;
Best Local Similarity 50.4%; Pred. No. 2 de-84;
Matches 240; Conservative 84; Mismatches 149; Indels 3; Gaps 3;

QY 22 TGLFINNKVFVSKNKTFEVINSPREELCHIVGCPHUUVEEAVQAADRAFSNCS WNCI 80
DB 32 TKIFINNHESKSKKFAFNPSTREQICVVRGCHUKPIVRAKAAQVAFQGSWPKL 91
QY 81 DPIDCKALYPLAELEQCKIVIASIFETLNGKA-ISSSPGVULVINYLKSSAGADKI 139
DB 92 DALSGRLLHULALIVKRAIALALETMTGKPFTHAFFIDLEGCIPTLRYFAGWADKI 151
QY 140 DPMIDGTHFSTYKQPLGVCCGLIPWNPFLIMMAWKIAPALVIGNTVVLKTAESTPL 199
DB 152 QCKTPTDNNVCTFHEPFGVCAITPWNPFLIMVWKIAPALCGNTWVTKPAQTPL 211
QY 200 SALYVSKYIPGAGITPGVINIVSGKLVVEATINHKIKKVAFTSTAGRIHYOSAA- 258
DB 212 TALYLGSLIKAGFPGVNVIVGFGTVAATSSHPQINKIAFTGSTEVCKLVKKAASR 271
QY 259 AGLKVTLELGGKSNIVAFAPAEKLVKAVQNTILGTYVNSGEVTCAGSPVVEESIVDKFI 318
DB 272 SNLKRVTELOCKNFCVCAADLDLAVEAHQGVFFNCGCCCTAASPRVFEQVYSEFV 331
QY 379 FIKPTVFGVADKMKLVKEEFSPVIVIKPKSADEVINMANUSEYGLAAGIHTSNINTA 438
DB 392 FIKPTVSEVTONMPLAKEEFPGVQPLIKPKSTEEVIKKNSTIDVGLTAAVETKNLKA 451
QY 439 LKVADEVNAGTVWNTYNIPHHAVPPGSPNASISIGRMSVIALQNYLQVAKVAPKL 494
DB 452 LKLASALESGIWINCYNALLYAQAPFGGFKMSGNGRELDYALAEYTEVKTIVIKL 507

RESULT 6
US-09-344-882-20
; Sequence 20, Application US/09344882
; Patent No. US20030162137A1
; GENERAL INFORMATION:
; APPLICANT: Nikolau, Basil J
; APPLICANT: Wurtele, Eve S
; APPLICANT: Oliver, David J
; APPLICANT: Behal, Robert
; APPLICANT: Schuabie, Patrick S
; APPLICANT: Ke, Jinshan
; APPLICANT: Johnson, Jerry L
; APPLICANT: Ailred, Carolyn C
; APPLICANT: Fatland, Beth

```

```

: APPLICANT: Lutziger, Isabella,
: APPLICANT: Wen, Tsui-Jung
: TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
: TITLE OF INVENTION: Acetyl CoA Levels in Plants
: FILE REFERENCE: 201573
: CURRENT APPLICATION NUMBER: US/09/344,882
: CURRENT FILING DATE: 1999-06-25
: PRIOR APPLICATION NUMBER: US 60/090,717
: PRIOR FILING DATE: 1998-06-26
: NUMBER OF SEQ ID NOS: 38
: SOFTWARE: PatentIn Ver. 2.2
: SEQ ID NO 20
: LENGTH: 538
: TYPE: PRT
: ORGANISM: Arabidopsis Thaliana
: US-09-344-882-20

```

Query Match	47.18	Score 1201	DB 9	Length 538
Best Local Similarity	48.94	Pred. No. 5e-83		
Matches	233	Conservative 82	Mismatches 159	Indels 2
Gaps	2			
QY	22	TGLFINNKFVPSKUNKTFEVINI:STEEELCHIVEGREDDUVEAVAUAKFNSGNSWNCID	81	
DB	58	TOLLINGNFVDSASGKTFPLDPTFGTEVIAHVAEGDAEDINRAVKAARTATDEGPWPKMS	117	
QY	62	PIDRCKALYRLAEILIEDOKDVIASETLDNGKAISS- RGDVDLVINYLKSSAFADKID	140	
DB	118	AYERSRVLLRFADIVKXHSBELASLETWONGKPYOQSITAEIPMFARLFRYYAGWAQKIH	177	
QY	141	GRMIDTGRTHSYTKRQPLGVCQGIIPWNEPILLMAWKIPALVTGNTVVLTKTAESTPLS	200	
DB	178	GLTIPADGNYOVHTLHEPIGVACQGIIPWNEPILLMFANKVGPALACNTIVLTKTEQPLT	237	
QY	201	ALYVSKYIPQAGIIPGVINIVSFGKIVVEVAIINHPKIKKVAFTGSTATGRHIYQSAA-A	259	
DB	238	AFYAKCLEAGLPGPVINIVSFGATAGAALASHMVDKLAFTGSTDTGKRVILGLAANS	297	
QY	260	GLKKVTLLELGKSPNIVFAADAELEKKAVQNIILGLIYVNSGEVCCAGSRVYVEESLYDFKIE	319	
DB	298	NLKPVTLELGKSPFIVFEDADDIKAVELAHFALFNQGCCAGSRITFVHKHYDFEVE	357	
QY	320	EFAKASES:IKVGDPFDESTFQGAQTSOMQJNKLILKYVDIGKNEGATLITGGERSGKGYF	379	
DB	358	KSKARALKRVYVGDPRFKGIIEQGPQIDLKQFEKVMKYIKSGIFESNATLECGGDOIADKGYF	417	
QY	380	IKPTVEGVIVKEDMKIVKEEITFGQPVTVTKFSADAEVINMANUSEYGLAAGIHTSNIHTAL	439	
DB	418	IQPTVFSNVKDDMLIAQDEILFIPVQSTILKFSVDDEVIKRANETRYGLAAGVFTKNLOTAN	477	
QY	440	KVADRVNAGTVWINTYNDFFHAYVPPGFGNAGLGHREMSVDALQNTYLVKAVRAKID	495	
DB	478	RYSRALRAGTVWNCDFVDDAIIIPPGCYKMSGNREKGIYSUNNYLOIKAVTVALN	533	

```

RESULT 7
US-10-293-865-20
: Sequence 20, Application US/10493865
: Publication No. US20030106090A1
: GENERAL INFORMATION:
: APPLICANT: Nikolau, Basil J
: APPLICANT: Wurtele, Eve S
: APPLICANT: Oliver, David J
: APPLICANT: Behal, Robert
: APPLICANT: Schnable, Patrick S
: APPLICANT: Ke, Jinshan
: APPLICANT: Johnson, Jerry L
: APPLICANT: Allred, Carolyn C
: APPLICANT: Fatland, Beth
: APPLICANT: Lotzger, Isabelle
: APPLICANT: Wen, Tsui-Jung
: TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
: TITLE OF INVENTION: Acetyl CoA Levels in Plants
: FILE REFERENCE: 217113

```

```

: CURRENT APPLICATION NUMBER: US2002/0293,865
: CURRENT FILING DATE: 2002-11-14
: PRIOR APPLICATION NUMBER: US 09/444,882
: PRIOR FILING DATE: 1999-06-25
: PRIOR APPLICATION NUMBER: US 60/090,717
: PRIOR FILING DATE: 1998-06-26
: NUMBER OF SEQ ID NOS: 48
: SOFTWARE: Patent In Ver. 3.1
: SEQ ID NO: 20
: LENGTH: 538
: TYPE: prt
: ORGANISM: Arabidopsis Thaliana
: 25:10-293-865-20

```

[illegible]

```

RESULT 8
US-09-344-882-24
: Sequence 24, Application US/09344882
: Patent No. US20020162137A1
: GENERAL INFORMATION:
: APPLICANT: Nikolau, Basil J
: APPLICANT: Wurtele, Eve S
: APPLICANT: Oliver, David J
: APPLICANT: Behal, Robert
: APPLICANT: Schmale, Patrick S
: APPLICANT: Ko, Jinsuan
: APPLICANT: Johnson, Jerry L
: APPLICANT: Alfred, Carolyn C
: APPLICANT: Fatland, Beth
: APPLICANT: Latziger, Isabelle
: APPLICANT: Wen, Tsui-Jung
: TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
: TITLE OF INVENTION: Acetyl CoA Levels in Plants
: FILE REFERENCE: 201573
: CURRENT APPLICATION NUMBER: US/09/344,882
: CURRENT FILING DATE: 1999-06-25
: PRIOR APPLICATION NUMBER: US 60/090,717

```


; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.2
; SEQ ID NO 24
; LENGTH: 534
; TYPE: PRT
; ORGANISM: Arabidopsis Thaliana
US-09-344-882-24

Query Match 45.98; Score 1169.5; DB 9; Length 534;
Best Local Similarity 48.5%; Pred. No. 1.2e-80;
Matches 238; Conservative 77; Mismatches 167; Indels 9; Gaps 3;
QY 6 MIVPIKIPNGLEYEQPTGLFINNKFPVSKONKTFEIVINPSTEEIECHIVEGREDDVEAV 65
DB 45 ITPPVKV-----EHTQLLGGFRFVDAVSGKFTPLDPNGEVIAQVSGDAEDYNRAV 97
QY 66 QAAAFASNSWNGTIDPDRKALYRLAELIEQDKVIASTETLDNCKAI-SSSRGVDVL 124
DB 98 AAARKAFDEGFWPKMTAYERSKILFFADLLIEKHNEIAALFTWENKPYEQSAQIEVPM 157
QY 125 VINYLKSSAGFADKIDGRMIDTGTHTPSYTKPQPGVCGQIIPWNPFLLMWAKIAPALV 184
DB 158 LARVFRYYAGWADKIHGMTPGDPGPHHVQTLHEPIGVAGQIIPWNPFLMLSKLGPALA 217
QY 185 TGNIVLKTAEPTPLSALYVSKYIPOAGIPPGVINIVSGPKIVVRAITNHPKIKKVAFT 244
DB 218 CGNIVVLKTAETPLSALYVSKYIPOAGIPPGVINIVSGPKIVVRAITNHPKIKKVAFT 277
QY 245 GSTATGRHIYQSA-AGLKVKVTLDELGSKSPNIVFADAEIKKAVQNIILGIYVNSGEVCCA 303
DB 278 GSTDVGKTIILASKSNLKVTLLEESHSFVCEADADQAVELAHAFALFFNQGCCCA 337
QY 304 GSRVTVVERSIYDKTEEFKAAESIYKVDPEDESTFOGAQTSOMQLNKILKYIDIGKNEG 363
DB 338 GSRTFVHRVYDEVKAKALPNVGDPPFKSGIEGQPOVDSEQFNKILKYIKHGEVAG 397
QY 364 ATLITGDRGSKGYFIKPTVFGDVKEDMRIVKBEIPGPVTVTKFSADEVINMANDSE 423
DB 398 ATLQAGDRGSKGYFIKPTVFGDVKEDMRIVKBEIPGPVTVTKFSADEVINMANDSE 457
QY 424 YGLAAGTHSNITALKVAQVNAQIVWNIYLFHHAVVFGGPNASGLGRKMSVDIALON 483
DB 458 YGLAAGVFTQNLDTAHLRALRVGIVWLNCFDVLASIPFGYKMSGIGREKGIYSLNN 517
QY 484 YLOVKAVRAKL 494
DB 518 YLOVKAVVTSL 528

RESULT 9
US-10-293-865-24
; Sequence 24, Application US/10293855
; Publication No. US20030106090A1
; GENERAL INFORMATION:
; APPLICANT: Nikolau, Basil J
; APPLICANT: Wurtele, Eve S
; APPLICANT: Oliver, David J
; APPLICANT: Behal, Robert
; APPLICANT: Schnable, Patrick S
; APPLICANT: Ke, Jinsan
; APPLICANT: Johnson, Jerry L
; APPLICANT: Allred, Carolyn C
; APPLICANT: Fatland, Beth
; APPLICANT: Lutziger, Isabelle
; APPLICANT: Wen, Tsui-Jung
; TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
; FILE REFERENCE: 217113
; CURRENT APPLICATION NUMBER: US/10/293,865
; CURRENT FILING DATE: 2002-11-13
; PRIOR FILING DATE: 1998-06-26
; PRIOR FILING DATE: 1998-06-26

; PRIOR APPLICATION NUMBER: US 60/060,717
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 24
; LENGTH: 534
; TYPE: PRT
; ORGANISM: Arabidopsis Thaliana
US-10-293-865-24

Query Match 45.98; Score 1169.5; DB 9; Length 534;
Best Local Similarity 48.5%; Pred. No. 1.2e-80;
Matches 238; Conservative 77; Mismatches 167; Indels 9; Gaps 3;
QY 6 MIVPIKIPNGLEYEQPTGLFINNKFPVSKONKTFEIVINPSTEEIECHIVEGREDDVEAV 65
DB 45 ITPPVKV-----EHTQLLGGFRFVDAVSGKFTPLDPNGEVIAQVSGDAEDYNRAV 97
QY 66 QAAAFASNSWNGTIDPDRKALYRLAELIEQDKVIASTETLDNCKAI-SSSRGVDVL 124
DB 98 AAARKAFDEGFWPKMTAYERSKILFFADLLIEKHNEIAALFTWENKPYEQSAQIEVPM 157
QY 125 VINYLKSSAGFADKIDGRMIDTGTHTPSYTKPQPGVCGQIIPWNPFLLMWAKIAPALV 184
DB 158 LARVFRYYAGWADKIHGMTPGDPGPHHVQTLHEPIGVAGQIIPWNPFLMLSKLGPALA 217
QY 185 TGNIVLKTAEPTPLSALYVSKYIPOAGIPPGVINIVSGPKIVVRAITNHPKIKKVAFT 244
DB 218 CGNIVVLKTAETPLSALYVSKYIPOAGIPPGVINIVSGPKIVVRAITNHPKIKKVAFT 277
QY 245 GSTATGRHIYQSA-AGLKVKVTLDELGSKSPNIVFADAEIKKAVQNIILGIYVNSGEVCCA 303
DB 278 GSTDVGKTIILASKSNLKVTLLEESHSFVCEADADQAVELAHAFALFFNQGCCCA 337
QY 304 GSRVTVVERSIYDKTEEFKAAESIYKVDPEDESTFOGAQTSOMQLNKILKYIDIGKNEG 363
DB 338 GSRTFVHRVYDEVKAKALRVNGDPPFKSGIEGQPOVDSEQFNKILKYIKHGEVAG 397
QY 364 ATLITGDRGSKGYFIKPTVFGDVKEDMRIVKBEIPGPVTVTKFSADEVINMANDSE 423
DB 398 ATLQAGDRGSKGYFIKPTVFGDVKEDMRIVKBEIPGPVTVTKFSADEVINMANDSE 457
QY 424 YGLAAGTHSNITALKVAQVNAQIVWNIYLFHHAVVFGGPNASGLGRKMSVDIALON 483
DB 458 YGLAAGVFTQNLDTAHLRALRVGIVWLNCFDVLASIPFGYKMSGIGREKGIYSLNN 517
QY 484 YLOVKAVRAKL 494
DB 518 YLOVKAVVTSL 528

RESULT 10
US 09 344 882-23
; Sequence 22, Application US/09344882
; Patent No. US20030137A1
; GENERAL INFORMATION:
; APPLICANT: Nikolau, Basil J
; APPLICANT: Wurtele, Eve S
; APPLICANT: Oliver, David J
; APPLICANT: Behal, Robert
; APPLICANT: Schnable, Patrick S
; APPLICANT: Ke, Jinsan
; APPLICANT: Johnson, Jerry L
; APPLICANT: Allred, Carolyn C
; APPLICANT: Fatland, Beth
; APPLICANT: Lutziger, Isabelle
; APPLICANT: Wen, Tsui-Jung
; TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
; FILE REFERENCE: 201573
; CURRENT APPLICATION NUMBER: US/09/344,882
; CURRENT FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 60/060,717

```

; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patent In Ver. 2.2
; SEQ ID NO 22
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Arabidopsis Thaliana
US-09-344-883-22

```

Query Match	45.0%	Score 1146.5	DB 9	Length 501
Best Local Similarity	47.1%	Pred. No. 6 2a-7a		
Matches	228	Conservative 95	Mismatches 154	Indels 7
Gaps	5			
QY	10	IKLPNCLEVEQPTGLFINNKFPVSKQNTKEVINFSIEEIEICHIEGKREDDVEEAVQAAD	69	
Db	12	VKLPE-IRF---TKLFINQGFIDAQSKTETIDPRNGEVIAITACGDKEDVDIAVNAAP	67	
QY	70	RAFSNGSNGIDPIDRGKALRYLAELEQDKDVIASTETLDNGKAISSSR-GVDVLVINY	128	
Db	68	YAFDHGPWRMTGFERAKLINFADLIEENIEELAKLDVDDGKLFQLGKYADIPATAGH	127	
QY	129	LKSSAGFADKIDGRMID-TGRTHFSYTKRPLGVCGQIIPWNPPLLMWAKIAPALVTGN	187	
Db	128	FRYNAGAADKIHGETLKMTRQSLFGYTLKEPIGVVGWNIIPWNPSPIMFATKVPAMAAGC	187	
QY	188	TVVLATAESTPLSALYSKYIPOAGIPPGVINIVSGFGKIVVEAITHHPKIKKVAFTGST	247	
Db	188	TWVWPAGQTSLSALFYAHLKSKEAGIPDGVINIVTGFSGTAGAAJASHMDVDVKSFTGST	247	
QY	248	ATGRHIVQSAAA-GLKKVTLLELGKSKSNIFVADAEALKKAVONIILGLIYVNSGEVFCAGSR	306	
Db	248	DWGRKIMQAAAAANLKKVSLGLGKSKSPLLIFNUADIDKAADLALGCFYNGKEICVASSR	307	
QY	307	VYVESRISYDKFIEEFKAASRISIKVGDHFDRESTDQGAQTSOMQLNKILKYVDITCKNEGATL	366	
Db	308	VFVQEGIYDKVVEKLEVEKAKLWTVDGDFDSTAKGQVUVUKQFEXILSYIEHCKNEGATL	367	
QY	367	ITGGERIGSKGYFFTKPTVFGDVEDMKRIVKEEIFGPVWTVTKFSKADEVINMANDSEYGL	426	
Db	368	LTCGKKAIGDKGYFTOPTIFADVTEDMKIYDEITFGPVMSLMKFKTVEEGIKCANNTKYGL	427	
QY	427	AAIGHTSINTALKVADRVNAGTWVINTYNDFIHHAVPFGGFNAGSLGRMSVUALQNLQYLQ	486	
Db	428	AAGILSQDIDLINTVSRISIKAGIIVWNCYCFGLDQCPYGYKMSGNCRESGMDALDNYLQ	487	
QY	487	VKAV 490		
Db	488	TKSV 491		

RESULT 11
US-100-293-865-22
Sequence 22, Application US/10293865
Publication No. US20030106040A1
GENERAL INFORMATION:
APPLICANT: Nikolae, Basil J
APPLICANT: Murtale, Eve S
APPLICANT: Oliver, David J
APPLICANT: Behal, Robert
APPLICANT: Schnable, Patrick S
APPLICANT: Ke, Jinshan
APPLICANT: Johnson, Terry L
APPLICANT: Allred, Carolyn C
APPLICANT: Fatland, Beth
APPLICANT: Lutziager, Isabelle
APPLICANT: Wen, Tsz-Jung
TITLE OF INVENTION: Materials and Methods for
Preparation of Acetyl CoA Level
FILE REFERENCE: 217113
CURRENT APPLICATION NUMBER: US/10/293865
CURRENT FILING DATE: 2002-11-13
PRIOR APPLICATION NUMBER: US 09/144
PRIOR FILING DATE: 1999-06-25

```

; PRIOR APPLICATION NUMBER: US 60/000,717
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patent In Ver. 3.1

```

Query Match	45.0%	Score	1146.5	DB	9	Length	5012
Best Local Similarity	47.1%	Prod.	No. 6,26-79				
Matches	228	Conservative	95	Mismatches	154	Indels	7
Gaps	5						
QY	10	IKLPNGIETTPQTLINNATVPKQKNTKEPVNISTEEPTETTYEYEGEGEPPFVQAAU	69				
Db	12	VKLPE-IKPE---TKLFGNGPTDAASCKTETTDTPRNGEVIAITAGDKEDVHIAVNAAR	67				
QY	79	PAFSNCSNWCITPITPDKATVPAAETPEKPKWIASFTITNKAKSSRPGRVPIVINY	128				
Db	68	YAFDHGPIWRIHGFERAKLNNKFAULFEENIEELAKLDVADGCKLQULKYALITATAGH	127				
QY	125	LESSAGFAFKIDGPMIDTQPTHESTYKPGGLTWQVQITWNPPIIMWAMPADAPVIND	187				
Db	128	FRYNAGAAADKIHGETLTKMTQSLGFCYTLKEFPIVWGNITIPWNPSTIMFATKVPAAAKA	187				
QY	186	IIVLTAFSTIPISALYSKYITPACITPGGVINIVASDEKIVVAFATINBETKEKVAFTST	247				
Db	188	IMVVPAGTQTSALITAFHASKKAGITPGGVINIVTGGSTAGAAASHMIVKVSFTST	247				
QY	246	ALGPIHYISAAA--GLEFVITETPGKSNIVFAIAIEKKAQVNIETATYNSGEVATASR	306				
Db	248	IVGKRLMGAAAAANLAKVSLDEPKSLITENWATIDRAALALDPTFKNGEIPVASSR	307				
QY	307	VYVESISTYKRTETTERAASSTPAGVQETSTETPQAGTSCMGNPTETAYVIGKNNGAIL	366				
Db	308	VFWQGLYDKVVEKIVFEKAKITWTWPTSTAGPQVAVKPKQVETKISYIEERKNEGAIL	367				
QY	367	ITGSEKPLDSKSYFTEKIVPGLVKKLDMIVPEFTPGVIVIKKSALEVINMANISEGIL	426				
Db	368	ITGSKAIGDKGYFTPTFAIVTIDMKLYQDPTGVMSIMPKETIVPEPTKANNIKYGL	427				
QY	427	AAAGHTSINTALVADVRVNGATVWITNYNIDFHAVFETGFKASGLGEMSLATQNYLQ	486				
Db	428	AAGLISQDILDIVNSGSIKAGLITWNGCYFGEFLMLTPQNYKMSNCTRESMEALINYLQ	487				
QY	487	VKAV 490					
Db	488	TKSV 491					

```

RESULT 12
US-10-268-518-4
: Sequence 4, Application US/10268518
: Publication: No. US2004010044A1
: GENERAL INFORMATION:
: APPLICANT: Hunter, John, Joseph
: TITLE OF INVENTION: 9136, A HUMAN ALDEHYDE DEHYDROGENASE
: TITLE OF INVENTION: FAMILY MEMBER AND USES THEREFOR
: FILE REFERENCE: MP101-214PIRM
: CURRENT APPLICATION NUMBER: US/10268,518
: CURRENT FILING DATE: 2002.10.10
: PRIOR APPLICATION NUMBER: 60/429,899
: PRIOR FILING DATE: 2001.10.16
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID No 4
: LENGTH: 492
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Consensus sequence
: US-10-268-518-4

```

```

Query Match      41.88; Score 1067; DR 9; Length 492;
Best Local Similarity 46.98; Pred No 6, file 74;
Matches 231; Conservative 98; Mismatches 136; Indels 28; Gaps 15;

QY 29 KFPVSKNKTETVNPSTEEELIC-HIYESGREDUVEEAVQAADKAPFNGSGWNGIDPI-DRG 86
DB 1 EWDUSASCKTFFVNNPANKGEVIGRVPEATAEDUADAANKAAKFAFKSGPWAKVPASEKA 60
QY 87 KALYELAELEIQQDKVLIASITELLNGKAISSSGRVEL--VINYLKSSAPPAIKING--R 142
DB 61 RTIPKIALDEBEREDELAALETIDLGKPLAEAKGDTFVGPFAIDEIHYAGWAPKLMGRRP 120
QY 143 MIDT-----GRTHFSYTRQPIGVGGQIIPWNEFLMKWANKIAPALVTRNTVVLKTAESTP 198
DB 121 VPLSLADGDEELNYIKREFLGVGVSISPNWFLPILAIWKLAIAIAGANTIVVIKPSQGP 180
QY 199 LSALYVSKYIPQAG---IPGVNIVSGFGKIVVEALINHPKIKKVAFTJSTATGPHIVQ 255
DB 181 LTALLAELEIEAGANNLPKGVNVVPGFAEVGQALLSHPDIDKISFTGSTEVGKLIME 240
QY 256 SAAA-GLKKVILELGGSPNIVFADAEIKKAVANILIGIYNSGFWCCAGSPVYVEESIY 314
DB 241 AAAAKNLKKVILELGGSPNIVFADAEIKKAVANILIGIYNSGFWCCAGSPVYVEESIY 300
QY 315 DKFEEFKAAASETIK-VSDPPEST-FQSAQISQMLNKLIL-KYVDIGKNEGATLITGGE 371
DB 301 DEFVEKLEKPKVKIKIISQPIISLQNIYGPILSGQGFQVWSYIFGKKEEAKVLCGGE 360
QY 373 LSSK-----NYFTPTVFEVKEIMPVKKEEIPSPVVIVTKRKALEVINMANDESYHIA 427
DB 361 DESKPIFGGYYVQPTITFTVTPMUKIMKEEIPSPVLEPLIKPKDULDEALIELANDTEYGLA 420
QY 428 AGIHTSNI-NTALKVADRVNAGTIVWYNTYNGFHHA---VPEGGE-NASGLGEMSVD-A 481
DB 421 AAYVETKIDILAKAEKVAKALEAIVWVNVIV-CVHAAPQIPPGGVKQSSGIGPEHGSKYIG 474
QY 482 QNYLQVKAVERAKL 494
DB 480 LEEYTEIKTVIRL 493

RESULT 13
US-10-175-696-21
; Sequence 21, Application US/10175696
; Patent No. US20010092659A1
; GENERAL INFORMATION:
; APPLICANT: Glücksmann, Maria Alexandra
; APPLICANT: Rudolph-Owen, Laura A.
; TITLE OF INVENTION: NOVEL HUMAN ENZYME FAMILY MEMBERS AND USES THEREOF
; FILE REFERENCE: 10448-193001
; CURRENT APPLICATION NUMBER: US/10175,696
; CURRENT FILING DATE: 2002-06-20
; PRIOR APPLICATION NUMBER: 10/067,658
; PRIOR FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: 60/266,140
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 09/823,901
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: PCT/US01/10720
; PRIOR FILING DATE: 2001-04-02
; PRIOR APPLICATION NUMBER: 60/193,920
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/862,658
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: PCT/US01/16380
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: 60/205,675
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: 09/882,837
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: PCT/US01/19319
;

; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/211,727
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 493
; TYPE: PPT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: cons-nus sequence
us-10-175-696-21

Query Match      41.48; Score 1056.5; DR 9; Length 493;
Best Local Similarity 46.88; Pred No. 4, file 72;
Matches 231; Conservative 98; Mismatches 136; Indels 29; Gaps 16;

QY 29 KFPVSKNKTETVNPSTEEELIC-HIYESGREDUVEEAVQAADKAPFNGSGWNGIDPI-DRG 86
DB 1 EWDUSASCKTFFVNNPANKGEVIGRVPEATAEDUADAANKAAKFAFKSGPWAKVPASEKA 60
QY 87 KALYELAELEIQQDKVLIASITELLNGKAISSSGRVEL--VINYLKSSAPPAIKING--R 142
DB 61 RTIPKIALDEBEREDELAALETIDLGKPLAEAKGDTFVGPFAIDEIHYAGWAPKLMGRRP 120
QY 143 MIDT-----GRTHFSYTRQPIGVGGQIIPWNEFLMKWANKIAPALVTRNTVVLKTAESTP 198
DB 121 VPLSLADGDEELNYIKREFLGVGVSISPNWFLPILAIWKLAIAIAGANTIVVIKPSQGP 180
QY 199 LSALYVSKYIPQAG---IPGVNIVSGFGKIVVEALINHPKIKKVAFTJSTATGPHIVQ 255
DB 181 LTALLAELEIEAGANNLPKGVNVVPGFAEVGQALLSHPDIDKISFTGSTEVGKLIME 240
QY 256 SAAA-GLKKVILELGGSPNIVFADAEIKKAVANILIGIYNSGFWCCAGSPVYVEESIY 314
DB 241 AAAAKNLKKVILELGGSPNIVFADAEIKKAVANILIGIYNSGFWCCAGSPVYVEESIY 300
QY 315 DKFEEFKAAASETIK-VSDPPEST-FQSAQISQMLNKLIL-KYVDIGKNEGATLITGGE 371
DB 301 DEFVEKLEKPKVKIKIISQPIISLQNIYGPILSGQGFQVWSYIFGKKEEAKVLCGGE 360
QY 372 PLGSK-----GYFTKPTVFEVKEIMPVKKEEIPSPVVIVTKRKALEVINMANDESYGL 426
DB 361 PDESKEVIGGGYYVQPTITFTVTPMUKIMKEEIPSPVLEPLIKPKDULDEALIELANDTEYGL 420
QY 427 AAGIHTSNI-NTALKVADRVNAGTIVWYNTYNGFHHA---VPEGGE-NASGLGEMSVD-A 480
DB 421 AAYVETKIDILAPAPPVAKALFAGIVWVNVIV-CVHAAPQIPPGGVKQSSGIGPEHGSKYIG 474
QY 481 QNYLQVKAVERAKL 494
DB 480 LEEYTEIKTVIRL 493

RESULT 14
US-09-823-901-9
; Sequence 9, Application US/09823901
; Patent No. US09001807A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: 21509 AND 33770, NOVEL HUMAN
; TITLE OF INVENTION: DEHYDROGENASES AND USES THEREOF
; FILE REFERENCE: 10448-036001
; CURRENT APPLICATION NUMBER: US/09/823,901
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/862,658
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: PCT/US01/16380
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: 60/205,675
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: 09/882,837
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: PCT/US01/19319
;

```

```

; FEATURE:
; OTHER INFORMATION: consensus sequence
US-09-823-901-9

Query Match      41.4%; Score 1056.5; DB 10; Length 493;
Best Local Similarity 46.8%; Pred No. 4 1e-72;
Matches 231; Conservative 98; Mismatches 136; Indels 29; Gaps 16;

QY 29 KFYPSKONKTEVINSTEELIC-HIYEGRELDVEAVAAALAKAFNSNWSNGIDPI-DKG 86
Db 1 EWDSASGKTFEVNPNANGGEVIGKVPATEADVAAVKAAKAFKSGPWWAKVPASPPA 60

QY 87 KALYRLAELEIQKQKQVIASTETLDNGKATISSRGDVL--VINYLKSSACFADKIDG--R 142
Db 61 RILRLADLEEPDELALETLDLGKPLAFKAGTEVCRALDELRYVYAGWAPKLMERP 120

QY 143 MIDT----GRTHFSYTKRQPLGCGQIIPWNEPFLMAMWAKIAPALVTGNTVVLKTAESTP 198
Db 121 VIPSLATDGEELNTRREPLGVGVISPNWPEPLLLALWKLAPALAAAGNTVVLKPSQTP 180

QY 199 LSALYYSKYIPQAG---IPPGVINIVSGFGKTVVFAITNHPKIKKVAFTGSTATCPHYQ 255
Db 181 LTFALLAELEBAGANNLPKGVNVVPGFGEVGGALLSHPDIDKISFTGSTEVKILME 240

QY 256 SAAA-GLKKVTLRLGCKSPNIFADAEALKKAVONTIILGIYNSGEVCCAGSRVYVERSIY 314
Db 241 AAAAKNLKVTLELGGKSPVIVFDDADLLKAVERTVFAFGNAQVCIAPSPKLLVHESIY 300

QY 315 DKFIEFKAESIK-VGPPFEST-FQCAQTSQMLNKII-KYVDIGKNEGATLITGGE 371
Db 301 DEVEKLKERVKKLIGDPSDNTNIGPLISEQGFQFVMSYEDGKPKGAKVLCGGE 360

QY 372 FLASK--GYFKPTVGLVKELMPVKKELFIPVVIVIKESAGEVINMANSEYGL 426
Db 361 RDSKEYLGSGYVQPTIFTDTVPDKIMKEIFGVPVPIIKPDLDEAIELANDTEYGL 420

QY 427 AAGIHTSNI-NITALKVADVNAVAGTVWINTYNDFHHA---VPEGGF-NASGLGREGMSVD-A 480
Db 421 AAVFTKDLARAFRAVKALEAGIVWVNDV-CVHAAPQLPFGGVKQSSGIGREHGKYG 479

QY 481 LONYLOVKAVRAKL 494
Db 480 LEYTEIKTVTKL 493

RESULT 15
US-09-815-242-10550
; Sequence 10550, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA-011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,408
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10550
; LENGTH: 496
; TYPE: PPT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10550

Query Match      40.2%; Score 1025; DB 10; Length 496;
Best Local Similarity 43.6%; Pred No. 1e-69;
Matches 206; Conservative 87; Mismatches 171; Indels 8; Gaps 4;

QY 24 LFINNKVPSKONKTEVINSTEELICHYEGRELDVEEAVAAADAFNSNWSNGIDPI 83
Db 16 LYINGFWPTGSGNKKMIASVNSDEKIAFEVATINAIWHAFAAQAFAQ 1WKDVIVV 73

QY 84 DWCKALYRLAELEIEGKQKQVIASTETLDNGKATISSSP--DWLVINYLKSSACFADKIDG 142
Db 74 TRSNLLKLTADLIEENQEHILAMVETLDNGKPIREFOSIDVPAASAIHFFYFASVIRGEES 133

QY 143 MIDTGRTHFSYTKRQPLGCGQIIPWNEPFLMAMWAKIAPALVTGNTVVLKTAESTP 202
Db 134 VKFEFKDILSVVKEPIGVVGGIIPWNEPFLMAMWAKIAPALAAVNTVVIHSSSTLSLL 193

QY 203 YYSKYIPQAGIIPPGVINIVSGFGKTVVFAITNHPKIKKVAFTGSTATCPHYQ 252
Db 194 ELFKPTVGLVKELMPVKKELFIPVVIVIKESAGEVINMANSEYGL 422

QY 253 PATLELGCKSANIPEDANWELKALGCVOLGILFNGQCVVACSVIVQSGIYQVGVIALK 412
Db 323 AASESIKVGDPDESTFQCAQTSQMLNKII-KYVDIGKNEGATLITGGERAG--SKGY 478
Db 313 EKFEQVNVGFPWEKDVEMCAQINEHQLEELIKYVEIGVKKCATLITGQRETNELKGA 472

QY 379 FKPTVFGDVKEDMKRIVKEEIPGPVTVTKPSADEVINMANSEYGLAAGIHTSINTA 448
Db 373 FLAPTLANGTNTMCVAQEFPGVPVATVIRKPEFEVIRLANDSEYGLGAVFSQDINVA 432

QY 439 LKVADVNAVAGTVWINTYNDFHHAIVPFGFNAGSLGREGMSVDIALQNYLQVAKV 490
Db 433 LKVAAGVPLGPGWVNTYNGLAAGLPGGKSGIPLHUKSMIDAVYTGKNT 484

Search completed: June 24, 2003, 10:47:03
Job time : 24.9275 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:11:00 : Search time 174.237 seconds
(without alignments)
1780 862 Million cell updates/sec

Title: US-09-830-751-2
Perfect score: 2550
Sequence: 1 MSHLPMTVPKLNGLNGLEYEQ... MSVFALNLYLQKAVPAKILK 495

Scoring table: PROSNM62

Gapop 10 0 , Gapext 0.5

Searched: 4569144 seqs, 6447 x 110 residues 4569144

Total number of hits satisfying chosen param: 0

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 100%
Listing first 45 summaries

Database		Pending_Patents_AA_Main	
1:	/cgn2_6/ptodata/1/paa/US06_COMB.pep.*	1:	/cgn2_6/ptodata/1/paa/US06_COMB.pep.*
2:	/cgn2_6/ptodata/1/paa/US07_COMB.pep.*	2:	/cgn2_6/ptodata/1/paa/US07_COMB.pep.*
3:	/cgn2_6/ptodata/1/paa/US08_COMB.pep.*	3:	/cgn2_6/ptodata/1/paa/US08_COMB.pep.*
4:	/cgn2_6/ptodata/1/paa/US09_COMB.pep.*	4:	/cgn2_6/ptodata/1/paa/US09_COMB.pep.*
5:	/cgn2_6/ptodata/1/paa/US10_COMB.pep.*	5:	/cgn2_6/ptodata/1/paa/US10_COMB.pep.*
6:	/cgn2_6/ptodata/1/paa/US11_COMB.pep.*	6:	/cgn2_6/ptodata/1/paa/US11_COMB.pep.*
7:	/cgn2_6/ptodata/1/paa/US12_COMB.pep.*	7:	/cgn2_6/ptodata/1/paa/US12_COMB.pep.*
8:	/cgn2_6/ptodata/1/paa/US13_COMB.pep.*	8:	/cgn2_6/ptodata/1/paa/US13_COMB.pep.*
9:	/cgn2_6/ptodata/1/paa/US14_COMB.pep.*	9:	/cgn2_6/ptodata/1/paa/US14_COMB.pep.*
10:	/cgn2_6/ptodata/1/paa/US15_COMB.pep.*	10:	/cgn2_6/ptodata/1/paa/US15_COMB.pep.*
11:	/cgn2_6/ptodata/1/paa/US16_COMB.pep.*	11:	/cgn2_6/ptodata/1/paa/US16_COMB.pep.*
12:	/cgn2_6/ptodata/1/paa/US17_COMB.pep.*	12:	/cgn2_6/ptodata/1/paa/US17_COMB.pep.*
13:	/cgn2_6/ptodata/1/paa/US18_COMB.pep.*	13:	/cgn2_6/ptodata/1/paa/US18_COMB.pep.*
14:	/cgn2_6/ptodata/1/paa/US19_COMB.pep.*	14:	/cgn2_6/ptodata/1/paa/US19_COMB.pep.*
15:	/cgn2_6/ptodata/1/paa/US20_COMB.pep.*	15:	/cgn2_6/ptodata/1/paa/US20_COMB.pep.*
16:	/cgn2_6/ptodata/1/paa/US21_COMB.pep.*	16:	/cgn2_6/ptodata/1/paa/US21_COMB.pep.*
17:	/cgn2_6/ptodata/1/paa/US22_COMB.pep.*	17:	/cgn2_6/ptodata/1/paa/US22_COMB.pep.*
18:	/cgn2_6/ptodata/1/paa/US23_COMB.pep.*	18:	/cgn2_6/ptodata/1/paa/US23_COMB.pep.*
19:	/cgn2_6/ptodata/1/paa/US24_COMB.pep.*	19:	/cgn2_6/ptodata/1/paa/US24_COMB.pep.*
20:	/cgn2_6/ptodata/1/paa/US25_COMB.pep.*	20:	/cgn2_6/ptodata/1/paa/US25_COMB.pep.*
21:	/cgn2_6/ptodata/1/paa/US26_COMB.pep.*	21:	/cgn2_6/ptodata/1/paa/US26_COMB.pep.*
22:	/cgn2_6/ptodata/1/paa/US27_COMB.pep.*	22:	/cgn2_6/ptodata/1/paa/US27_COMB.pep.*
23:	/cgn2_6/ptodata/1/paa/US28_COMB.pep.*	23:	/cgn2_6/ptodata/1/paa/US28_COMB.pep.*
24:	/cgn2_6/ptodata/1/paa/US29_COMB.pep.*	24:	/cgn2_6/ptodata/1/paa/US29_COMB.pep.*
25:	/cgn2_6/ptodata/1/paa/US30_COMB.pep.*	25:	/cgn2_6/ptodata/1/paa/US30_COMB.pep.*
26:	/cgn2_6/ptodata/1/paa/US31_COMB.pep.*	26:	/cgn2_6/ptodata/1/paa/US31_COMB.pep.*
27:	/cgn2_6/ptodata/1/paa/US32_COMB.pep.*	27:	/cgn2_6/ptodata/1/paa/US32_COMB.pep.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2538	94.5	514	21	US-09-791-537-47094, A
2	2538	94.5	514	27	US-60-360-039-22438, A
3	1807	70.9	518	21	US-09-791-537-38066, A
4	1728.5	67.8	520	27	US-09-791-537-19395, A
5	1728.5	67.8	520	21	US-60-360-039-17499, A
6	1696	64.5	514	21	US-09-791-537-70112, A

7	1621.5	63.6	519	25	US-10-179-131-8039, A
8	1455	57.1	447	21	US-09-791-537-24771, A
9	1452	56.9	509	21	US-09-791-537-61775, A
10	1452	56.9	509	27	US-60-360-039-22495, A
11	1443.5	56.6	501	21	US-09-791-537-70619, A
12	1432.5	56.2	498	27	US-60-360-039-4846, A
13	1412.5	55.4	497	10	US-08-612-853-11, A
14	1412.5	55.4	497	21	US-09-791-537-73644, A
15	1399	54.9	496	10	US-08-612-853-2, A
16	1399	54.9	496	21	US-09-791-537-9158, A
17	1399	54.9	496	22	US-09-847-2081-59, A
18	1399	54.9	496	22	US-09-847-208B-59, A
19	1394	54.7	496	10	US-08-612-853-3, A
20	1394	54.7	497	10	US-08-612-853-10, A
21	1362.5	53.4	494	16	US-08-612-853-2, A
22	1362.5	53.4	494	16	US-09-791-537-26770, A
23	1362.5	53.4	495	22	US-09-847-208-11, A
24	1362.5	53.4	495	22	US-09-847-208B-11, A
25	1362.5	53.4	496	10	US-08-612-853-9, A
26	1362.5	53.4	503	21	US-09-791-537-115973, A
27	1325.5	52.3	497	11	US-09-791-537-21613, A
28	1297	50.9	382	16	US-09-248-746-17019, A
29	1297	50.9	382	27	US-60-360-039-17019, A
30	1276.5	50.0	519	21	US-09-791-537-148084, A
31	1274.5	50.0	519	21	US-09-791-537-74080, A
32	1266	49.5	539	25	US-10-179-141-5408, A
33	1261.5	49.5	500	21	US-09-791-537-64443, A
34	1256.5	49.4	517	27	US-60-360-039-1227, A
35	1256.5	49.3	517	27	US-60-412-418-1227, A
36	1254.5	49.2	532	27	US-60-360-039-12695, A
37	1253.5	49.2	517	19	US-09-538-092-851, A
38	1253.5	49.2	517	21	US-09-791-537-74079, A
39	1253.5	49.2	517	27	US-60-389-987-1242, A
40	1253.5	49.2	517	27	US-60-412-418-1242, A
41	1252.5	49.1	495	21	US-09-791-537-4937, A
42	1252.5	49.1	495	27	US-60-360-039-23162, A
43	1252.5	49.1	517	21	US-09-791-537-60432, A
44	1251.5	49.1	494	21	US-09-791-537-66667, A
45	1250.5	49.0	517	21	US-09-791-537-71773, A

ALIGNMENTS

RESULT 1
US-09-791-537-47094, Application US/09791537
Sequence 47094, Applicant Biocomix, Inc.
GENERAL INFORMATION:
APPLICANT: Biocomix, Inc.
APPLICANT: Leber, Joseph
TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
TITLE OF INVENTION: METHODS OF USE THEREOF
FILE REFERENCE: 261/210
CURRENT APPLICATION NUMBER: US/09/791-537
CURRENT FILING DATE: 2001 02 22
NUMBER OF SEQ IN NOS: 154055
SOFTWARE: Patent in version 3.0
SEQ ID NO 47094
LENGTH: 519
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-791-537-47094

Query Match	94.5%	Score	2538	DB ID	21	Length	519
Best Local Similarity	99.88%	Pred. No	5e-230				
Matches	493	Conservative	0	Mismatches	1	Indels	0
Gaps	0						
Q1	2	SHLPMVPIKLNGLNGLEYEQPTGLFINNKVFPSKNTFEVINFTSTEEETHIYEGEDVV	61				
Q2	25	SHLPMVPIKLNGLNGLEYEQPTGLFINNKVFPSKNTFEVINFTSTEEETHIYEGEDVV	84				
Q3	42	SHLPMVPIKLNGLNGLEYEQPTGLFINNKVFPSKNTFEVINFTSTEEETHIYEGEDVV	121				

Db 85 EEAVQAARAFNSWNGIDPDRGKALYPLAFI IQPKDVIASITELUNGKAISSSGD 144
 QY 122 VDLVINYLKSSAGFADKIDGRMIDTGRTHFSYTKRQPLGVCQGIIPWNPFLMMAWKIAP 181
 Db 145 VDLVINYLKSSAGFADKIDGRMIDTGRTHFSYTKRQPLGVCQGIIPWNPFLMMAWKIAP 204
 QY 182 ALVTGNTVVLKTAESTPLSALYSKYIPOAGIPPGVINIVSGFGKIYVEAITNHPKIKKV 241
 Db 205 ALVTGNTVVLKTAESTPLSALYSKYIPOAGIPPGVINIVSGFGKIYVEAITNHPKIKKV 264
 QY 242 AFTGSTATGRHLYQSAAGLKKVYLEGGKSNIVFADAEELKAVQNIILGIIYNSGVC 301
 Db 265 AFTGSTATGRHLYQSAAGLKKVYLEGGKSNIVFADAEELKAVQNIILGIIYNSGVC 324
 QY 302 CAGSRVYVESIYDKFIEEFKAASESIKVGDPDEFTFOGATOSOMQLNKILKYVDIGKN 361
 Db 325 CAGSRVYVESIYDKFIEEFKAASESIKVGDPDEFTFOGATOSOMQLNKILKYVDIGKN 384
 QY 362 EGATLTGGRLGSKGYFIKPTVFGDVKEDMRIVKKEIEFGPVVTITKPKSADEVINMAND 421
 Db 385 EGATLTGGRLGSKGYFIKPTVFGDVKEDMRIVKKEIEFGPVVTITKPKSADEVINMAND 444
 QY 422 SEYGLAAGIHTSNINTALKVADRVNAGTVMINTYDFHRAVPPGGFNASGLGRENMSVDAL 481
 Db 445 SEYGLAAGIHTSNINTALKVADRVNAGTVMINTYDFHRAVPPGGFNASGLGRENMSVDAL 504
 QY 482 QNYLOVKAVRAKLD 495
 Db 505 QNYLOVKAVRAKLD 518

RESULT 2

US-60-360-039-22438
 ; Sequence 22438, Application US/60360039
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yungwei
 ; APPLICANT: Chen, Xianfeng
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)A
 ; CURRENT APPLICATION NUMBER: US/60/360,039
 ; CURRENT FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 22438
 ; LENGTH: 519
 ; TYPE: PRT
 ; ORGANISM: Saccharomyces cerevisiae
 US-60-360-039-22438

Query Match 99.8%, Score 2538, PR 27, Length 513,
 Best Local Similarity 99.8%, Pred. No. 5e-230,
 Matches 493, Conservative 0, Mismatches 1, Indels 0, Gaps 0:
 QY 2 SHLPMTVFIKPLNGLEYEQTGLFINNKEVPSKQNKTEFVNPSTEEETICHYEGREDDV 61
 Db 25 SHLPMTVFIKPLNGLEYEQTGLFINNKEVPSKQNKTEFVNPSTEEETICHYEGREDDV 84
 QY 62 EEAVQAARAFNSWNGIDPDRGKALYPLAFI IQPKDVIASITELUNGKAISSSGD 121
 Db 85 EEAVQAARAFNSWNGIDPDRGKALYPLAFI IQPKDVIASITELUNGKAISSSGD 144
 QY 122 VDLVINYLKSSAGFADKIDGRMIDTGRTHFSYTKRQPLGVCQGIIPWNPFLMMAWKIAP 181
 Db 145 VDLVINYLKSSAGFADKIDGRMIDTGRTHFSYTKRQPLGVCQGIIPWNPFLMMAWKIAP 204
 QY 182 ALVTGNTVVLKTAESTPLSALYSKYIPOAGIPPGVINIVSGFGKIYVEAITNHPKIKKV 241
 Db 205 ALVTGNTVVLKTAESTPLSALYSKYIPOAGIPPGVINIVSGFGKIYVEAITNHPKIKKV 264

QY 242 AFTGSTATGRHLYQSAAGLKKVYLEGGKSNIVFADAEELKAVQNIILGIIYNSGVC 301
 Db 265 AFTGSTATGRHLYQSAAGLKKVYLEGGKSNIVFADAEELKAVQNIILGIIYNSGVC 324
 QY 302 CAGSRVYVESIYDKFIEEFKAASESIKVGDPDEFTFOGATOSOMQLNKILKYVDIGKN 361
 Db 325 CAGSRVYVESIYDKFIEEFKAASESIKVGDPDEFTFOGATOSOMQLNKILKYVDIGKN 384
 QY 362 EGATLTGGRLGSKGYFIKPTVFGDVKEDMRIVKKEIEFGPVVTITKPKSADEVINMAND 421
 Db 385 EGATLTGGRLGSKGYFIKPTVFGDVKEDMRIVKKEIEFGPVVTITKPKSADEVINMAND 444
 QY 422 SEYGLAAGIHTSNINTALKVADRVNAGTVMINTYDFHRAVPPGGFNASGLGRENMSVDAL 481
 Db 445 SEYGLAAGIHTSNINTALKVADRVNAGTVMINTYDFHRAVPPGGFNASGLGRENMSVDAL 504
 QY 482 QNYLOVKAVRAKLD 495
 Db 505 QNYLOVKAVRAKLD 518

RESULT 3

US-09-791-537-48066
 ; Sequence 38066, Application US/09791537
 ; GENERAL INFORMATION:
 ; APPLICANT: Bionomix, Inc.
 ; APPLICANT: Debo, Derek
 ; APPLICANT: Danzer, Joseph
 ; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
 ; FILE REFERENCE: 261/210
 ; CURRENT APPLICATION NUMBER: US/09/791,537
 ; CURRENT FILING DATE: 2001-02-22
 ; NUMBER OF SEQ ID NOS: 153055
 ; SOFTWARE: Patent In version 4.0
 ; SEQ ID NO 48066
 ; LENGTH: 518
 ; TYPE: PRT
 ; ORGANISM: Pichia angusta
 US-09-791-537-48066

Query Match 70.9%, Score 1807, PR 21, Length 518,
 Best Local Similarity 68.2%, Pred. No. 5.9e-161,
 Matches 337, Conservative 72, Mismatches 81, Indels 2, Gaps 2:
 QY 2 SHLPMTVFIKPLNGLEYEQTGLFINNKEVPSKQNKTEFVNPSTEEETICHYEGREDDV 61
 Db 22 SHLPMTVFIKPLNGLEYEQTGLFINNKEVPSKQNKTEFVNPSTEEETICHYEGREDDV 81
 QY 62 EEAVQAARAFNSWNGIDPDRGKALYPLAFI IQPKDVIASITELUNGKAISSSGD 121
 Db 82 EEAVQAARAFNSWNGIDPDRGKALYPLAFI IQPKDVIASITELUNGKAISSSGD 140
 QY 122 VDLVINYLKSSAGFADKIDGRMIDTGRTHFSYTKRQPLGVCQGIIPWNPFLMMAWKIAP 181
 Db 141 VDLVINYLKSSAGFADKIDGRMIDTGRTHFSYTKRQPLGVCQGIIPWNPFLMMAWKIAP 200
 QY 182 ALVTGNTVVLKTAESTPLSALYSKYIPOAGIPPGVINIVSGFGKIYVEAITNHPKIKKV 241
 Db 201 ALVTGNTVVLKTAESTPLSALYSKYIPOAGIPPGVINIVSGFGKIYVEAITNHPKIKKV 260
 QY 242 AFTGSTATGRHLYQSAAGLKKVYLEGGKSNIVFADAEELKAVQNIILGIIYNSGVC 300
 Db 261 AFTGSTATGRHLYQSAAGLKKVYLEGGKSNIVFADAEELKAVQNIILGIIYNSGVC 320
 QY 302 CAGSRVYVESIYDKFIEEFKAASESIKVGDPDEFTFOGATOSOMQLNKILKYVDIGKN 360
 Db 321 CAGSRVYVESIYDKFIEEFKAASESIKVGDPDEFTFOGATOSOMQLNKILKYVDIGKN 380
 QY 362 EGATLTGGRLGSKGYFIKPTVFGDVKEDMRIVKKEIEFGPVVTITKPKSADEVINMAND 420
 Db 481 EGATLTGGRLGSKGYFIKPTVFGDVKEDMRIVKKEIEFGPVVTITKPKSADEVINMAND 440

SEQ ID NO 70612
 LENGTH: 519
 TYPE: PR1
 ORGANISM: Saccharomyces cerevisiae
 US-09-791-537-70612

Query Match 66.5% Score 1695; DB 21; Length 519;
 Best Local Similarity 63.7%; Pred. No. 2.3e-150;
 Matches 316; Conservative 80; Mismatches 94; Indels 6; Gaps 2;

QY 2 SHLPMTVPIKPLNGLEYEPTGLFINNKFVPSKONKTFEVIINPSTEEETICHYEGREDEVE 61
 DB 24 SOAPLRVPITLPGNFTVEPTGLFTNGEFVASKQKTFDVIINPNEEKITTVKAMEDDV 83
 QY 62 EEAQVADRAFSNSW--NGIDPIDRGKALYRLAELEQUKUVIASIEILDNKAISSSK 119
 DB 84 DEAAVALKLLKRSVYCRAGV---RAKALFNLAIDIVEKHQETFLAAIESMDNCKSLFCAR 139
 QY 120 GDVVLVINYLKSSAGFADKIDGRMIDTGRTHFSYTKRPIGVCGGQIIPWNPFLMMWAKI 179
 DB 140 GDVALVSKYLRSCGWDKTIYGNVIDTGKNEFTYSYKEPLGVCGQIIPWNPFLMMWAKI 199
 QY 180 APALVTGNTVVLKTAESTPLSALYVSKYIPQAGIPPGVINIVSGFGKIYVEAITNHPKTK 239
 DB 200 GPALATGNTVVLKPAETTPLSALFASQLCQAEAGIPAGVNNILPGSGKVVGERLSAHPDVK 259
 QY 240 KVAFTGSTATGRHYGSAAGLKKVILELQSKSPNIVEALAEELKRAVQNIILGIYNSGE 299
 DB 260 KIATGSTATGRHLMKVAADTVKVTLEIGKSPNIVFADADUDKRAVKNIAFGIFYNSGE 319
 QY 300 VCCAGSRVYVESIYDKFIEEFKAASESIKVGDPDPDESTFOCAQTSOMOLNKLKYVDIG 359
 DB 320 VCCAGSRVYIQDTVYEEVLQKDYTESLKVGPDPDEEVFOGAQTSKQLHKILDYDVA 379
 QY 360 KNEGATLITGERLGSGKGYFIKTVFGDVKEDMRIYKEEIFGPVTVTKFSADAEVINMA 419
 DB 380 KSEGARLVTCGARHSGKGYFVKTVFADVKEDMRIYKEEIVFGPIVTVSKFSTVDEVIAMA 439
 QY 420 NDSEYGLAAGIHTSNITALKVADRVNAGTVWINTYNDHFHVAVPFGGPNASGLGRMSVD 479
 DB 440 NDSQYGLAAGIHTINDINKAVDSKRVKAKTVWINTYNNHFNQNVFGGPNASGLGRMSVEA 499
 QY 480 ALONYLOKAVRAKLD 495
 DB 500 ALSNYTQTSVRIAD 515

RESULT 7
 US-10-179-131-8239
 ; Sequence 8239, Application US/10179131
 ; GENERAL INFORMATION:
 ; APPLICANT: HARE, ROBERTA S.
 ; APPLICANT: SHAW, KAREN J.
 ; APPLICANT: SHIMER JR., GEORGE H.
 ; APPLICANT: KESSLER, MARCO
 ; APPLICANT: NOLLING, JORK
 ; APPLICANT: ZENG, JIANDONG
 ; APPLICANT: GREENE, JONATHAN R.
 ; TITLE OF INVENTION: CANDIDA ALBICANS NUCLEIC ACIDS AND POLYPEPTIDES,
 ; FILE OF INVENTION: AND USES THEREOF
 ; FILE REFERENCE: 2976-4031
 ; CURRENT APPLICATION NUMBER: US/10/179,131
 ; CURRENT FILING DATE: 2002-06-21
 ; NUMBER OF SEQ ID NOS: 10194
 ; SEQ ID NO 8239
 ; LENGTH: 519
 ; TYPE: PR1
 ; ORGANISM: Candida albicans
 US-10-179-131-8239

Query Match 63.6% Score 1621.5; DB 25; Length 515;
 Best Local Similarity 61.5%; Pred. No. 2e-143;
 Matches 303; Conservative 86; Mismatches 101; Indels 3; Gaps 3;

QY 4 LPMTVPITKPLNGLEYEPTGLFINNKFVPSKONKTFEVIINPSTEEETICHYEGREDEVE 63
 DB 26 LPLVSKLTITPKGIIYNGPLGLEFINNKHKKKIDFEVSPSDEETITVYEALDEHDIPT 85
 QY 64 AVQAALRAFSNSWNGIDPIDRGKALYRLAELEQUKUVIASIEILDNKAISSSKVD 124
 DB 86 AVEAAQAYING-WAQGPPEORSKVIPLKLADEENAEELIAQETWNGKSKSNAKCHVA 144
 QY 124 LVINYLKSSAGFADKIDGRMIDTGRTHFSYTKRPIGVCGGQIIPWNPFLMMWAKIATPAL 183
 DB 145 LTAAYFSCGWDKTIIGSQTNGTNTFNHYTORVPL-VCGQIIPWNPFLMMWAKIAPVL 203
 QY 184 VTGNIVVLKIAESTPLSALYVSKYIPQAGIPPGVINIVSGFGKIYVEAITNHPKAKVAF 243
 DB 204 ATSTTVEKTAESTPLSALYVSKYIPQAGIPPGVINIVSGFGKIYVEAITNHPKAKVAF 263
 QY 244 TGTSTATGRHYGSAAGLKKVILELQSKSPNIVEALAEELKRAVQNIILGIYNSGEV 402
 DB 264 TGTSTATGRHLMKLAATESNIKKVTLPLGKSPNIVFADADUDKRAVKNIAFGIFYNSGEV 423
 QY 303 AGSRVYVESIYDKFIEEFKAASESIKVGDPDPDESTFOCAQTSOMOLNKLKYVDIGKNE 462
 DB 324 AGSPLLTQSGVYIQDVVEKFEAAESKVGSPNPFDETFMGACQVSGVLSLKYVESKSG 483
 QY 363 GATLITGCEPFGSGKGYFIKTVFGDVKEDMRIYKEEIFGPVTVTKFSADAEVINMANUS 422
 DB 384 GATVTVTGARADKCGYFVKTVFADVKKIMDIWREELFGPVVTLKFDITVDEAVELANUS 443
 QY 423 EYGLAAGIHTSNITALKVADRVNAGTVWINTYNDHFHVAVPFGGPNASGLGRMSVDIAQ 482
 DB 444 EYGLAAGIHTADVNKKIDVANKRVKAGIWNVNTYNDHFHVAVPFGGPNASGLGRMSDEVIH 503
 QY 483 NYLOKAVRAKLD 495
 DB 504 EYTOVRAVRMKIN 516

RESULT 8
 US-09-791-537-26771
 ; Sequence 26771, Application US/09791537
 ; GENERAL INFORMATION:
 ; APPLICANT: Bionomix, Inc.
 ; APPLICANT: Debo, Derek
 ; APPLICANT: Danzer, Joseph
 ; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
 ; FILE OF INVENTION: METHODS OF USE THEREOF
 ; FILE REFERENCE: 261/210
 ; CURRENT APPLICATION NUMBER: US/09/791,537
 ; CURRENT FILING DATE: 2001-02-22
 ; NUMBER OF SEQ ID NOS: 153055
 ; SOFTWARE: Patent In version 4.0
 ; SEQ ID NO 26771
 ; LENGTH: 497
 ; TYPE: PR1
 ; ORGANISM: Aspergillus niger
 US-09-791-537-26771

Query Match 57.1% Score 1456; DB 21; Length 497;
 Best Local Similarity 56.0%; Pred. No. 1e-127;
 Matches 277; Conservative 84; Mismatches 141; Indels 4; Gaps 4;
 QY 1 MSHPMTVEIKPLNGLEYEPTGLFINNKFVPSKONKTFEVIINPSTEEETICHYEGREDE 60
 DB 1 MSDLEAT ITTPNGVKYEQPLGLFDTGFEVCAEGKTFETINPSNEKPIVAVHEATEKD 98
 QY 61 VEEAVQAADRAFSNSWNGIDPIDRGKALYRLAELEQUKUVIASIEILDNKAISSSRG 120
 DB 59 VDTAVAAAKRAF-EGSWKQVITSTGRKMLIKLADIFERDAELASLEADNCKSTIMAHG 117
 QY 121 DVEGLVINYLKSSAGFADKIDGRMIDTGRTHFSYTKRPIGVCGGQIIPWNPFLMMWAKI 180
 DB 118 DIAGAAGLPYVGWAKIRGQUTIDINSETINVTYBETFGVGGIIPWNPFLMMWAKIP 177

QY 181 PALVTGNTVVLKATSLPASAIVYSSKYTPQAGLTPGVJNIVSVSRKIVWFAITNHPKIKK 240
 DB 178 PALATGNTVVLKATSLPASAIVYSSKYTPQAGLTPGVJNIVSVSRKIVWFAITNHPKIKK 240
 QY 241 VAFSTGATGRHIGYQAA-AGIKKVTLELGGKSNIVFAUAELEKAVQNIILGIIYNSGE 299
 DB 238 VAFSTGATGRHIGYQAA-AGIKKVTLELGGKSNIVFAUAELEKAVQNIILGIIYNSGE 299
 QY 300 VTCAGSRVYVERSTIVKFIETPKFAASFSIKVQDGFDPDSFPOSAGTSSQMLNKLKYVDI 459
 DB 298 VTCAGSRVYVERSTIVKFIETPKFAASFSIKVQDGFDPDSFPOSAGTSSQMLNKLKYVDI 459
 QY 360 KNEGATLITGGERLGSQVFTKPTVFGDVKEDMRIVKEIFGPVVTITKFKSADEVINMA 419
 DB 358 KIAGAIVAVGSPFPHSTGYFQLPIVFEGLVVISMKINLEELPSPVVIWQKIKVDELKIKG 417
 QY 420 NDEYGLAAGIHTSNINTALKVADRVNAGTIVWINTYDFHHAVPFGFNAGSLGREGMSVD 479
 DB 418 NSTSYGLAAGIHTKVTITAIRVSNALRAGIVWVNSYNLIQVQVPGFGKESGIGREGSY 477
 QY 480 ALONYLOVKAVPAKL 494
 DB 478 ALONYLOVKAVPAKL 492

RESULT 9

US-09-791-537-61775

Sequence 61775, Application US/09791537

GENERAL INFORMATION:

APPLICANT: Bionomix, Inc.

APPLICANT: Debe, Derek

APPLICANT: Danzer, Joseph

TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS

TITLE OF INVENTION: METHODS OF USE THEREOF

FILE REFERENCE: 261/210

CURRENT APPLICATION NUMBER: US/09/791,537

NUMBER OF SEQ ID NOS: 153055

SOFTWARE: PatentIn version 3.0

SEQ ID NO 61775

LENGTH: 500

TYPE: PRT

ORGANISM: Saccharomyces cerevisiae

US-09-791-537-61775

Query Match 56.9%, Score 1452, DB 21, Length 500,
 Best Local Similarity 54.8%, Pred. No. 2-127,
 Matches 272, Conservative 92, Mismatches 128, Indels 4, Gaps 3,

QY 3 HLPMTVPK--LPNGLEYEQPTGLFINNKVFESKONKIFEVINPSTEEELCHIVEGRED 60
 DB 5 HFDIAEPVKITLPLNGLTVEQPTGLFINNKFKAGDGKTYPVEDPSTENTVCEVSSATTED 64
 QY 61 VEEAVQAADPAFNSGWSNGITLIDPSKALYLAELIEQKIVIASIETLNGKALSSSRG 120
 DB 65 VEYATPCAPAFHETEMATQGPPEPGTISLKALELESQILVSSIEALDNGKTLALAP 124
 QY 121 DVLVTGNTVWLKTAESTPLSALYVSKYIPOAGIPPGVINIVSGFGKIIVFAITNHPKIKK 180
 DB 125 DVTIACLRLDAAYADKNGETINTGICYMNFTEPTICVCGQIIPWNEPIMLAWKIA 184
 QY 181 PALVTGNTVWLKTAESTPLSALYVSKYIPOAGIPPGVINIVSGFGKIIVFAITNHPKIKK 240
 DB 185 PALAMGNVCILKPAAVTPELNALYFASLCCKKVGIPAGVNVIVPGPGTIVCAALNDPPIK 244
 QY 241 VAFSTGATGRHIGYQAAAGLKVTLLELGGKSNIVFAUAELEKAVQNIILGIIYNSGE 299
 DB 245 LAFTGTEVGKSVAVDSSFSNLKKTITLGGKSAHLVFDANIKKTIPLNVNGIFKKNAG 304
 QY 300 VTCAGSRVYVERSTIVKFIETPKFAASFSIKVQDGFDPDSFPOSAGTSSQMLNKLKYVDI 459
 DB 298 VTCAGSRVYVERSTIVKFIETPKFAASFSIKVQDGFDPDSFPOSAGTSSQMLNKLKYVDI 459
 QY 360 KNEGATLITGGERLGSQVFTKPTVFGDVKEDMRIVKEIFGPVVTITKFKSADEVINMA 419
 DB 358 KIAGAIVAVGSPFPHSTGYFQLPIVFEGLVVISMKINLEELPSPVVIWQKIKVDELKIKG 417
 QY 420 NDEYGLAAGIHTSNINTALKVADRVNAGTIVWINTYDFHHAVPFGFNAGSLGREGMSVD 479
 DB 418 NSTSYGLAAGIHTKVTITAIRVSNALRAGIVWVNSYNLIQVQVPGFGKESGIGREGSY 477
 QY 480 ALONYLOVKAVPAKL 494
 DB 478 ALONYLOVKAVPAKL 492

QY 459 EKNEGATLITGGERLGSQVFTKPTVFGDVKEDMRIVKEIFGPVVTITKFKSADEVINMA 419
 DB 455 EKNEGATLITGGERLGSQVFTKPTVFGDVKEDMRIVKEIFGPVVTITKFKSADEVINMA 419
 QY 419 ANDEYGLAAGIHTSNINTALKVADRVNAGTIVWINTYDFHHAVPFGFNAGSLGREGMSV 478
 DB 425 ANSSEFGSGGTETESI STLKVKAKMLKAGIIVWINTYDFHHAVPFGFNAGSLGREGMSV 478
 QY 478 ALONYLOVKAVPAKL 494
 DB 485 EVYHAYTEVKAVPAKL 500

RESULT 10

US-09-791-537-61775

Sequence 22495, Application US/60360039

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Chen, Xianfeng

APPLICANT: Goldman, Barry S.

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)A

CURRENT APPLICATION NUMBER: US/60/360,039

CURRENT FILING DATE: 2002 02 21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 22495

LENGTH: 500

TYPE: PRT

ORGANISM: Saccharomyces cerevisiae

US-60-360-039-22495

Query Match 56.9%, Score 1452, DB 27, Length 500,

Best Local Similarity 54.8%, Pred. No. 2-127,

Matches 272, Conservative 92, Mismatches 128, Indels 4, Gaps 3,

QY 3 HLPMTVPK--LPNGLEYEQPTGLFINNKVFESKONKIFEVINPSTEEELCHIVEGRED 60
 DB 5 HFDIAEPVKITLPLNGLTVEQPTGLFINNKFKAGDGKTYPVEDPSTENTVCEVSSATTED 64
 QY 61 VEEAVQAADPAFNSGWSNGITLIDPSKALYLAELIEQKIVIASIETLNGKALSSSRG 120
 DB 65 VEYATPCAPAFHETEMATQGPPEPGTISLKALELESQILVSSIEALDNGKTLALAP 124
 QY 121 DVLVTGNTVWLKTAESTPLSALYVSKYIPOAGIPPGVINIVSGFGKIIVFAITNHPKIKK 180
 DB 125 DVTIACLRLDAAYADKNGETINTGICYMNFTEPTICVCGQIIPWNEPIMLAWKIA 184
 QY 181 PALVTGNTVWLKTAESTPLSALYVSKYIPOAGIPPGVINIVSGFGKIIVFAITNHPKIKK 240
 DB 185 PALAMGNVCILKPAAVTPELNALYFASLCCKKVGIPAGVNVIVPGPGTIVCAALNDPPIK 244
 QY 241 VAFSTGATGRHIGYQAAAGLKVTLLELGGKSNIVFAUAELEKAVQNIILGIIYNSGE 299
 DB 245 LAFTGTEVGKSVAVDSSFSNLKKTITLGGKSAHLVFDANIKKTIPLNVNGIFKKNAG 304
 QY 300 VTCAGSRVYVERSTIVKFIETPKFAASFSIKVQDGFDPDSFPOSAGTSSQMLNKLKYVDI 459
 DB 298 VTCAGSRVYVERSTIVKFIETPKFAASFSIKVQDGFDPDSFPOSAGTSSQMLNKLKYVDI 459
 QY 360 KNEGATLITGGERLGSQVFTKPTVFGDVKEDMRIVKEIFGPVVTITKFKSADEVINMA 419
 DB 358 KIAGAIVAVGSPFPHSTGYFQLPIVFEGLVVISMKINLEELPSPVVIWQKIKVDELKIKG 417
 QY 420 NDEYGLAAGIHTSNINTALKVADRVNAGTIVWINTYDFHHAVPFGFNAGSLGREGMSV 478
 DB 425 ANSSEFGSGGTETESI STLKVKAKMLKAGIIVWINTYDFHHAVPFGFNAGSLGREGMSV 478
 QY 478 ALONYLOVKAVPAKL 494
 DB 485 EVYHAYTEVKAVPAKL 500

Db 485 EVVHAYTEVAVRIKL 500

RESULT 11
US-09-791-537-70619
Sequence 70619, Application US/09791537
GENERAL INFORMATION:
APPLICANT: Bionomix, Inc.
APPLICANT: Debe, Derek
TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
FILE REFERENCE: 261/210
CURRENT APPLICATION NUMBER: US/09/791,537
CURRENT FILING DATE: 2001-02-22
SOFTWARE: PatentIn version 3.0
SEQ ID NO 70619
LENGTH: 501
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-791-537-70619

Query Match 56.6%; Score 1443.5; DB 21; Length 501;
Best Local Similarity 54.7%; Pred. No. 1,3e-126;
Matches 272; Conservative 92; Mismatches 128; Indels 5; Gaps 4;

QY 3 HLPWTPYIK--LPLNGLEYQPTGLFINNKVFPSKONKTFEVPINPSTPEEICHVYEGREDD 60
Db 5 HETAEVKITLPLNGTYEQPTGLFINNKFKADQKTYVEDPSTENTCEVSSATTE 64
QY 61 VEEVAQAADRAFSGNSWNGIDPIDRGKALRYLAELIEQDKVIASITELDKGKATS-SSR 119
Db 65 VEVAIECADRAFHDTEWATQDPRERGRLLSKLADELSQIDLVSSIEALDNGKTLAPKAR 124
QY 120 GDVDLVINLYKSSAGFADKIDGRMDTGRTHSYTKRQPLGVCGGQIIPWNPFLMWAKWI 179
Db 125 GDVTIACINCLRDAAAYADKVGRTINTDGYMNFITLEPVGCGGQIIPWNPFLMWAKWI 184
QY 180 APALVTGNTVLTAEESTPLSALYVSKYIPOAGIPPGVINIVSGFGKIVVEAINTNHPKIK 239
Db 185 APALMGNCVILKPAAVTFLNALYFASLCCKVGIAGVNVIVPGFRTVGAALNDPPIR 244
QY 240 KVAFTGSTATGRHI--YQSAAGHKKVITELGKSPNIVFADAEKAKVQNIILGIYNSG 298
Db 245 KLAFTGSTEVGKSVAVDSSESNLKKITILEGKSAHLVFDGANIKKTLPLNLVNGIFKNAG 304
QY 299 EVCCAGSRVYVEESIYDKFLIEBFKAASES-IKVGDPDEFSTFGAQTSMQNLKILKYVD 357
Db 305 QICSSSRIVYQGIYDELLAAFKAYLETEIKVGNPFDRANFQGAITNRQOFTIMNYID 364
QY 358 IGKNEGATLITGGERIAGSKGYFIKPTVFGDVKRMKIVKEETFGPVVTVTKFKSADEVIN 417
Db 365 IGKKEGAKILTGKGVKDGKGYFIRPTVYFVVDNEMRIYKEETFGPVTVAKFKTLEEGVE 424
QY 418 MANDSEYGLAAGHTSNINTALKVADRVNAGTVIWTNDFHVAVPFGFNAGSLGGRMS 477
Db 425 MANSSEFGSLGSGTETSLTGLKVAKMLKAGTVIWTNDFSDSRVPGCVKQSGYGRMG 484
QY 478 VDALKNTYLOVKAVRAKL 494
Db 485 EVVHAYTEVAVRIKL 501

RESULT 12
US-60-360-039-3846
Sequence 3846, Application US/60360039
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Chen, Xiantong
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.

Db 485 EVVHAYTEVAVRIKL 500

RESULT 13
US-08-612-853-11
Sequence 11, Application US/08612853
GENERAL INFORMATION:
APPLICANT: Achatz, Gernot
APPLICANT: Oberkofler, Hannes
APPLICANT: Simon, Birgit
APPLICANT: Unger, Andrea
APPLICANT: Lechenauer, Erich
APPLICANT: Hirschwehr, Reinhold
TITLE OF INVENTION: Recombinant Cladosporium Bertharum
NUMBER OF SEQUENCES: 117
CORRESPONDENCE ADDRESS:
ADDRESSEE: pemie & Edmonds LLP
STREET: 1155 Avenue of the Americas

QY 25 FINNKVPDSKONKTFEVPINPSTPEEICHVYEGREDDVFAVAAARAFNSNWSNGTIDPID 84
Db 1 FINNEVEGVGDKKTFEVPINPSTPEEICHVYEGREDDVFAVAAARAFNSNWSNGTIDPID 84
QY 85 RCKALRYLAELIEQDKVIASITELDKGKATS-SSR 119
Db 60 RGIYLLKLADELKLLAALAAVESLNGKSTIMAKEDGAVVGTIRYVGGWAKLEKNT 119
QY 145 DTGRTHSYTKRQPLGVCGGQIIPWNPFLMWAKWIAPALVGVNIVVVKIAETSDLSALYV 204
Db 120 DISPDSEFHTROEPLGVCGGQIIPWNPFLMWAKWIAPALVGVNIVVVKIAETSDLSALYV 179
QY 295 SKYIPVAGIPPGVINIVSGFGKIVVEAINTNHPKIKVAFGSLAIGPRTVYNSAA 264
Db 180 AQVKEAGHPGVLNIIISGFGRIAGAAASIMDIUKVAFSTWVGKIMKAAAFESNLKK 239
QY 264 VTILEDGKSPNIVFADAEKAKVQNIILGIYNSG 298
Db 240 VTILEDGKSPNIVFADAEKAKVQNIILGIYNSG 298
QY 324 ASHSTHVGDPDEFSTFGAQTSMQNLKILKYVD 357
Db 300 RAOONKVGDPDEFSTFGAQTSMQNLKILKYVD 357
QY 384 VFGVWKEKMKIVKEETFGPVVTVTKFKSADEVIN 417
Db 360 IFTNVHEDMKIMKEETFGPVVTVTKFKSADEVIN 417
QY 444 RVNAGIYVNI 494
Db 420 HLRAGTIVVNIYKTYCTVLEKVTIFRANONI 479
QY 481 LQNYLVKAVRAKL 495
Db 480 LANTOCKSVAIKLN 494

Query Match 56.2%; Score 1442.5; DB 27; Length 498;
Best Local Similarity 54.1%; Pred. No. 1.4e-126;
Matches 258; Conservative 88; Mismatches 114; Indels 25; Gaps 4;

QY 25 FINNKVPDSKONKTFEVPINPSTPEEICHVYEGREDDVFAVAAARAFNSNWSNGTIDPID 84
Db 1 FINNEVEGVGDKKTFEVPINPSTPEEICHVYEGREDDVFAVAAARAFNSNWSNGTIDPID 84
QY 85 RCKALRYLAELIEQDKVIASITELDKGKATS-SSR 119
Db 60 RGIYLLKLADELKLLAALAAVESLNGKSTIMAKEDGAVVGTIRYVGGWAKLEKNT 119
QY 145 DTGRTHSYTKRQPLGVCGGQIIPWNPFLMWAKWIAPALVGVNIVVVKIAETSDLSALYV 204
Db 120 DISPDSEFHTROEPLGVCGGQIIPWNPFLMWAKWIAPALVGVNIVVVKIAETSDLSALYV 179
QY 295 SKYIPVAGIPPGVINIVSGFGKIVVEAINTNHPKIKVAFGSLAIGPRTVYNSAA 264
Db 180 AQVKEAGHPGVLNIIISGFGRIAGAAASIMDIUKVAFSTWVGKIMKAAAFESNLKK 239
QY 264 VTILEDGKSPNIVFADAEKAKVQNIILGIYNSG 298
Db 240 VTILEDGKSPNIVFADAEKAKVQNIILGIYNSG 298
QY 324 ASHSTHVGDPDEFSTFGAQTSMQNLKILKYVD 357
Db 300 RAOONKVGDPDEFSTFGAQTSMQNLKILKYVD 357
QY 384 VFGVWKEKMKIVKEETFGPVVTVTKFKSADEVIN 417
Db 360 IFTNVHEDMKIMKEETFGPVVTVTKFKSADEVIN 417
QY 444 RVNAGIYVNI 494
Db 420 HLRAGTIVVNIYKTYCTVLEKVTIFRANONI 479
QY 481 LQNYLVKAVRAKL 495
Db 480 LANTOCKSVAIKLN 494

CITY: New York
 STATE: NY
 COUNTRY: USA
 ZIP: 10036-2711
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/612.853
 FILING DATE: 26-FEB-1996
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/AT94/00120
 FILING DATE: 02-MAR-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Friebe, Thomas E
 REGISTRATION NUMBER: 29,258
 REFERENCE/DOCKET NUMBER: 6530-021-999
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-790-9090
 TELEFAX: 212-869-9741
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 11:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 497 amino acids
 TYPE: amino acid
 STRANDEDNESS: unknown
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Aspergillus
 US-08-612-853-11

Query Match 55.4%, Score 1412.5, DB 10, Length 497;

Best Local Similarity 54.5%, Pred. No. 1e-123;
Matches 270, Conservative 88, Mismatches 132, Indels 5; Gaps 3;

```

QY 1 MSHLPMTVPKLPNGLEYEPTGLFINNKFVPSKQNTFEVINFSEIEEICHVEGPEDD 50
DB 1 MSDLFTTETPV---IKYEQPLGLFINNEFVKGVGKTFQVFNPSNEKVITSVHEATEKD 57

61 VEEAVQAAADRAFSNGSNGIDPIDPGKALYPLAELEFQKVKVIASITELDNGKAISSSPG 120
DB 58 VIVAVAAAPAAF-EGPWPQVTPSEPGIILINKLADLMERLIDILAAIESLUNGKAFMAKV 116

121 DVDLVINYLKSSAGFADKIDGRMTDGTGRTHSEYTKRQPLGVCGGQIIPWNEPFLIMWAKIA 180
DB 117 DLANSGLRYVAGWADKHGQITDINPELTYTHPEPVGVCGGQIIPWNEPFLIMWSWKIG 176

181 PALVTGNTVVLKTAESTPLSALYVSKYIPQAGIPPGVINIVSGFGKTVVPAITNHPKIRK 240
DB 177 PAVAAGNTVVLKTAQOTPLSALYAAKLKEAPFPAGVINIVSGFGRTAGAAISSHMDIUK 236

241 VAFVGTATGPHIYQSA--AGLKKVTELGKSNIVFAADAEKAVQNIILGIYVNSGE 299
DB 237 VAFVGTATGPHIYQSA--AGLKKVTELGKSNIVFAADAEKAVQNIILGIYVNSGE 296

300 VCCAGSRVYVESIVTKFTEFFKAASESTKVGLPFELESTFGAGTISQMLNKILKYVDIG 359
DB 297 CCCAGSPHIVQEGITYKFEVAREKFAKKNFVNPFEQTFQGPQVSGTQFDRIMEYINHG 356

360 KNGRATILITGGERIGSKGYFTKTPVGVGVREDMPVIVKEEIPGPVVTVTKESADEVINMA 419
DB 357 KKAGATVATGGRHNEGYFTQPTVFTDVTSDMKIAQEELFGPVVLIQKEKVAEALKIG 416

420 NUSEYGLAAGLHISNTALKAUVKAVNACIWIINTYNDPHHVPFGFGFNAGSLGREGMSVD 479
DB 417 NSTDYGAAAVHINKVNTAIRVSNALKAGTVIWNINNYNMIYSQAPFGFGKSGLGREGSY 476

480 ALONYLOVKAVRAKL 494
DB 477 ALENYTOIKTVHYRL 491
  
```

RESULT 15

US-08-612-853-2

; Sequence 2, Application US/08612853

; GENERAL INFORMATION:

; APPLICANT: Achatz, Gernot

; APPLICANT: Oberkofler, Hannes

; APPLICANT: Simon, Birgit

; APPLICANT: Unger, Andrea

; APPLICANT: Lechenauer, Erich

DB 477 ALENYTOIKTVHYRL 491

RESULT 14

US-09-791-537-73644

; Sequence 73644, Application US/09791537

; GENERAL INFORMATION:

; APPLICANT: Bionomix, Inc.

; APPLICANT: Debe, Derek

; APPLICANT: Danzer, Joseph

; TITLE OF INVENTION: THREE DIMENSIONAL SIMULATIONS OF PROTEIN FAMILIES AND FAMILY ME

; TITLE OF INVENTION: METHODS OF USE THEREOF

; FILE REFERENCE: 261/210

; CURRENT APPLICATION NUMBER: US/09/791,537

; CURRENT FILING DATE: 2001-02-22

; NUMBER OF SEQ ID NOS: 153055

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 73644

; LENGTH: 497

; TYPE: PRT

; ORGANISM: Emericella nidulans

US-09-791-537-73644

Query Match 55.4%, Score 1412.5, DB 21, Length 497;

Best Local Similarity 54.5%, Pred. No. 1e-123;

Matches 270, Conservative 88, Mismatches 132, Indels 5; Gaps 3;

```

QY 1 MSHLPMTVPKLPNGLEYEPTGLFINNKFVPSKQNTFEVINFSEIEEICHVEGPEDD 60
DB 1 MSDLFTTETPV---IKYEQPLGLFINNEFVKGVGKTFQVFNPSNEKVITSVHEATEKD 57

61 VEEAVQAAADRAFSNGSNGIDPIDPGKALYPLAELEFQKVKVIASITELDNGKAISSSPG 120
DB 58 VIVAVAAAPAAF-EGPWPQVTPSEPGIILINKLADLMERLIDILAAIESLUNGKAFMAKV 116

121 DVDLVINYLKSSAGFADKIDGRMTDGTGRTHSEYTKRQPLGVCGGQIIPWNEPFLIMWAKIA 180
DB 117 DLANSGLRYVAGWADKHGQITDINPELTYTHPEPVGVCGGQIIPWNEPFLIMWSWKIG 176

181 PALVTGNTVVLKTAESTPLSALYVSKYIPQAGIPPGVINIVSGFGKTVVPAITNHPKIRK 240
DB 177 PAVAAGNTVVLKTAQOTPLSALYAAKLKEAPFPAGVINIVSGFGRTAGAAISSHMDIUK 236

241 VAFVGTATGPHIYQSA--AGLKKVTELGKSNIVFAADAEKAVQNIILGIYVNSGE 299
DB 237 VAFVGTATGPHIYQSA--AGLKKVTELGKSNIVFAADAEKAVQNIILGIYVNSGE 296

300 VCCAGSRVYVESIVTKFTEFFKAASESTKVGLPFELESTFGAGTISQMLNKILKYVDIG 359
DB 297 CCCAGSPHIVQEGITYKFEVAREKFAKKNFVNPFEQTFQGPQVSGTQFDRIMEYINHG 356

360 KNGRATILITGGERIGSKGYFTKTPVGVGVREDMPVIVKEEIPGPVVTVTKESADEVINMA 419
DB 357 KKAGATVATGGRHNEGYFTQPTVFTDVTSDMKIAQEELFGPVVLIQKEKVAEALKIG 416

420 NUSEYGLAAGLHISNTALKAUVKAVNACIWIINTYNDPHHVPFGFGFNAGSLGREGMSVD 479
DB 417 NSTDYGAAAVHINKVNTAIRVSNALKAGTVIWNINNYNMIYSQAPFGFGKSGLGREGSY 476

480 ALONYLOVKAVRAKL 494
DB 477 ALENYTOIKTVHYRL 491
  
```

```

: APPLICANT: Hirschwehr, Reinhold
: TITLE OF INVENTION: Recombinant Cladosporium Herbarum
: NUMBER OF INVENTIONS: 1
: NUMBER OF SEQUENCES: 117
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fendle & Edmonds LLP
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: NY
: COUNTRY: USA
: ZIP: 10036-2711
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSeq Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/612,853
: FILING DATE: 26-FEB-1996
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/AT94/00120
: FILING DATE: 02-MAR-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Friebe, Thomas E
: REGISTRATION NUMBER: 29,258
: REFERENCE/LOCKET NUMBER: 6550-021-999
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-790-9090
: TELEFAX: 212-869-9741
: TELEX: 66141 PENNIE
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 496 amino acids
: TYPE: amino acid
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: internal
: ORIGINAL SOURCE:
: ORGANISM: Cladosporium herbarum
: DEVELOPMENTAL STAGE: Spores and vegetative hyphae
:
: US-08-612-853-2

```

```

Query Match      54.9%; Score 1399; DB 10; Length 496;
Best Local Similarity 53.8%; Pred. No. 26-122;
Matches 263; Conservative 89; Mismatches 135; Indels 2; Gaps 2;

QY 7 TVPIKLPNGLEYHPTGLFPIKKVPKSKONTFEVINPSTEEIEICHIEGREDDVEEAVO 66
Db 3 SVQLETPHSGKYEQPTGLFINNEPVKGGKTFDINPDSVITQVHEATEKDVDTAVA 62

QY 67 AADRAFSNCSWNCIDPIDRGKALYRLAELIEDKDVIASIETDLNGKAISSRGDVLVI 126
Db 63 AARQAF-PGSGWPIFTPFNPKILNLANLFEKNTDILLAAVESLDNGKATSMARVTSACAS 121

QY 127 NVLKSSAGFADKIDGRMIDTGRTHFSYTKROPGLGCGOIIPWNEPILLMWAKIAPALVTG 186
Db 122 GCLRYGGWADKLTGKVITDTPDTFNYYKKEPIGVCRSHLSLELPLLMWAKIGPAIACG 181

QY 187 NTVVLKATSTPLSALYSKYIPQAGIIPGVINIVSGKIVVEFATINHPKIKKVAFTGS 246
Db 182 NTVVLKATSTPLGGLVAASLVKEAGPPGVINIVSGFGKVGAGAAALSSHMDVDKVAFTGS 241

QY 247 TATGRHIYQSAAGLKKVVTLELGGKSPNIVFAELKKAQVNTIILGIYNSGEVCCAGS 305
Db 242 TVVGRITLKAASSNLKKVITELGGKSPNIVFEDADIDNAISWYVNGIFFNHGQCCAGS 301

QY 306 RVVVEESIYDKFIEFKAASESIVKGDFFDESTFGAQTOSMQLNKILKYVDIGKNEGAT 365
Db 302 RVVVEESIYDKFVQKFKRAQKNVVDPPFAALTFGPQVSKVQFDKIMETQACKDASAT 361

```

```

QY 456 LIIGGERLGSKYFIRPIVFGIWKELMKIVKKEIFGPVVIVIRKSADEVINMANISSEYG 425
Db 362 VETGSEKGRKYFIRPIVFGIWKELMKIVKKEIFGPVSTAKERIRKELAIKLNASTYIG 421

QY 426 LAAGIHTSNINIALKVADKVNAGTVMNINYNDFHVAPEGGENANGIGREMSVIALONYI 485
Db 422 LAAAVHTENLNTATEVSNALFKAGTVMNINYNILHEUMPPFGYKESIGPFGIATANYI 481

QY 486 QVKAVRAKL 494
Db 482 QTKTVSIRL 490

```

Search completed: June 24, 2003, 10:40:46
Job time : 182.207 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein protein search, using sw mod.i

Run on: June 24, 2003, 10:14:35 ; Search time 53.4876 Seconds
(without alignments)
2375.712 Million cell updates

Title: US-09-830-751-2

Perfect score:

Sequence: |MSHLPMTVPKIKLPNGLEVEI|MSVIAIUNYIIVKAVKAKID 495

Scoring table. BL0SUM62

Gapop 10.0 , Gapext

Searched: 1171708 seqs, 257189365 residues

Total number of hits satisfying chosen parameters: 1171708

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_AA_New: *

- 1: /cgn2_6/prodata/2/paa/pnt_NEW_01mk.pdf
- 2: /cgn2_6/prodata/2/paa/ins07_NEW_01mk.pdf
- 3: /cgn2_6/prodata/2/paa/ins07_NEW_01mk.pdf
- 4: /cgn2_6/prodata/2/paa/ins08_NEW_01mk.pdf
- 5: /cgn2_6/prodata/2/paa/ins08_NEW_01mk.pdf
- 6: /cgn2_6/prodata/2/paa/ins10_NEW_01mk.pdf
- 7: /cgn2_6/prodata/2/paa/ins10_NEW_01mk.pdf

Pred No is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query		Length	DB	ID	Description
	Score	Match				
1	4250	100%	495	5	US-09-4830-751-2	Sequence 2, Appl1
2	2538	99.5	519	6	US-10-360-493-32438	Sequence 23438, A
3	1728.5	67.8	520	6	US-10-360-493-1749	Sequence 1749, A
4	1452	56.9	500	6	US-10-369-493-22495	Sequence 22495, A
5	1432.5	56.2	498	6	US-10-369-493-1846	Sequence 4846, A
6	1274.5	50.0	519	6	US-10-219-051B-6969	Sequence 6969, A
7	1274.5	50.0	519	6	US-10-219-051B-6973	Sequence 6973, A
8	1254.5	49.2	532	6	US-10-369-493-12695	Sequence 12695, A
9	1253.5	49.2	517	6	US-10-319-051B-6241	Sequence 6241, A
10	1253.5	49.2	517	6	US-10-319-051B-6975	Sequence 6975, A
11	1253.5	49.2	530	5	US-09-449-016-10636	Sequence 10636, A
12	1252.5	49.1	495	6	US-10-369-493-33162	Sequence 33162, A
13	1248.5	48.8	516	6	US-10-369-493-4555	Sequence 4555, A
14	1239.5	48.6	514	7	US-60-453-135-14871	Sequence 14871, A
15	1239.5	48.6	514	7	US-60-453-050-14871	Sequence 14871, A
16	1239.5	48.6	514	7	US-60-466-412-14871	Sequence 14871, A
17	1236	48.5	544	6	US-10-369-493-5715	Sequence 5715, A
18	1234.5	48.4	552	6	US-10-425-114-59042	Sequence 59042, A
19	1231	48.3	532	5	US-09-949-015-9207	Sequence 9207, A
20	1230	48.2	518	5	US-09-724-676-64510	Sequence 64510, A
21	1230	48.2	518	5	US-09-724-676-64510	Sequence 64510, A
22	1226.5	48.1	552	6	US-10-425-114-49963	Sequence 49963, A
23	1226	48.1	549	6	US-10-425-114-51294	Sequence 51294, A
24	1225.5	48.1	500	5	US-09-830-751-4	Sequence 4, Appl1
25	1224	48.0	813	6	US-10-437-948-112958	Sequence 112958, A
26	1223.5	47.9	518	6	US-10-144-779-472	Sequence 472, A

ALIGNMENTS

RESULT 1

US-09-830-751-2

; Sequence 2, Application US/09830751

; GENERAL INFORMATION:

; APPLICANT: Suthers, Patrick F

; APPLICANT: Cameron, Douglas C.

TABLE IV
INVENTION OF THE INVENTOR

; TITLE OF INVENTION: Organisms

FILE REFERENCE: 44-38861-751

DIFFERENTIAL APPLIED POLYMER SYMPOSIA

PREVIOUS APPLICATION NUMBER: 611/151,440

FILED FILING DATE 1999-08-30

; PRIK APPLICATION NUMBER: PCT/US00/23878

; PRIOR FILING DATE: 2000-08-30

; NUMBER OF SEQ ID NOS: 23

; SOFTWARE: PA

; SEQ ID NO 2

```

; LENGTH: 495
; TYPE: DDT

```

ORGANISM: *Saccharomyces cerevisiae*

US-09-830-751-2

Quarry Match

Best Local Similarity	100.0%	Prod No 22-197
29.17	29.17	29.17

Matches	495	Conservative	0	Mismatches	0	Indels	0	Gaps	0
<p> </p>									

QY	1	MSHLPMTVPIKLPLNGLEYEQDTGLF	INNKFPVSKQNKTFEVINPSTFETICHVYKREDD	60
DB	1	MSHLPMTVPIKLPLNGLEYEQDTGLF	INNKFPVSKQNKTFEVINPSTFETICHVYKREDD	60
QY	61	VEFAVQAAMAFSSGNNINIDTPKAKALYPLAET	FEFQWVYASFTINCFKALSSSPG	120
DB	61	VEFAVQAAMAFSSGNNINIDTPKAKALYPLAET	FEFQWVYASFTINCFKALSSSPG	120
QY	121	DWLEVINYKSSAGFAIKIDGPMIDGTHFSYTKPQPVVQ	IGQIIPNNEPILMMAWKIA	180
DB	121	DWLEVINYKSSAGFAIKIDGPMIDGTHFSYTKPQPVVQ	IGQIIPNNEPILMMAWKIA	180
QY	181	PALVTCNTVVLTKAETSPISALVYSKYIQAGIPPRV	INVSQEGKIVVEAITNHPKK	240
DB	181	PALVTCNTVVLTKAETSPISALVYSKYIQAGIPPRV	INVSQEGKIVVEAITNHPKK	240
QY	241	VAFTGSTATGPIVQSAAMVILFKVTLELGKSNIVFA	IALKPAVQVILIGIIVYNSGEV	300
DB	241	VAFTGSTATGPIVQSAAMVILFKVTLELGKSNIVFA	IALKPAVQVILIGIIVYNSGEV	300
QY	301	CGASFPVYVESLYDKFIEPPFAASPSIKVTDHFD	FEFTFQZAFPSAMQINPTIKVYDTCP	360

Db 301 CCAGSRVYVEESIVDKFIEEFKAASESIKVGDPDFDEFTQGAUTSOMQLNKILKYVDICK 360
QY 361 NEGATLITGGERHASKGYFIKPTVFGVDKEDMPVKEEIPGPVVITKPKSADEVINMAN 420
Db 361 NEGATLITGGERHASKGYFIKPTVFGVDKEDMPVKEEIPGPVVITKPKSADEVINMAN 420
QY 421 DSEYGLAAGIHTSNINTALKVADRVNAGTWTWINTYNDPFHHAHPFGGFNASGLGREGMSVDA 480
Db 421 DSEYGLAAGIHTSNINTALKVADRVNAGTWTWINTYNDPFHHAHPFGGFNASGLGREGMSVDA 480
QY 481 LONYLOVKAVRAKLD 495
Db 481 LONYLOVKAVRAKLD 495

RESULT 2
US-10-369-493-22438
: Sequence 22438, Application US/10369493
: GENERAL INFORMATION:
: APPLICANT: Cao, Yongwei
: APPLICANT: Hinkle, Gregory J.
: APPLICANT: Slater, Steven C.
: APPLICANT: Goldman, Barry S.
: APPLICANT: Chen, Xianfeng
: TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
: FILE REFERENCE: 38-10(52052)B
: CURRENT APPLICATION NUMBER: US/10/369,493
: PRIOR FILING DATE: 2003-02-28
: PRIOR APPLICATION NUMBER: US 60/360,039
: NUMBER OF SEQ ID NOS: 47374
: SEQ ID NO 22438
: LENGTH: 519
: TYPE: PRT
: ORGANISM: Saccharomyces cerevisiae
US-10-369-493-22438

Query Match 99.5% Score 2538; DB 6; Length 519;
Best Local Similarity 99.8%; Pred. No. 2, 1e 196;
Matches 493; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 SHLPMTVPKLPNGLEVEQPTGLFINNKEVPVSKQNKTEFVINSPEEFIECHIEGREDV 61
Db 25 SHLPMTVPKLPNGLEVEQPTGLFINNKEVPVSKQNKTEFVINSPEEFIECHIEGREDV 84
QY 62 BEAVQAADRAFNSGWSNGIDPDRGKALYPLAELIEQDKDVIASITLDNGKATSSRGD 121
Db 85 BEAVQAADRAFNSGWSNGIDPDRGKALYPLAELIEQDKDVIASITLDNGKATSSRGD 144
QY 122 VDLVINYLKSSAGFADKIDGRMIDTGRTHFSYTKRQPLGVCGQIIPWNPFLMLMAWKIAP 181
Db 145 VDLVINYLKSSAGFADKIDGRMIDTGRTHFSYTKRQPLGVCGQIIPWNPFLMLMAWKIAP 204
QY 182 ALVTGNTVVLKTAESTPLSALYVSKYIPQAGIPGVINIVSGFGKIVVEFATNHPKIKV 241
Db 205 ALVTGNTVVLKTAESTPLSALYVSKYIPQAGIPGVINIVSGFGKIVVEFATNHPKIKV 264
QY 242 AFTGSTATGRHIYQSAAGLKKVTLLELGGKSNIVFADAEKLVAVONTILGIYNSGEVC 301
Db 265 AFTGSTATGRHIYQSAAGLKKVTLLELGGKSNIVFADAEKLVAVONTILGIYNSGEVC 324
QY 302 CAGSRVYVEESIVDKFIEEFKAASESIKVGDPDFDEFTQGAUTSOMQLNKILKYVDIGKN 361
Db 325 CAGSRVYVEESIVDKFIEEFKAASESIKVGDPDFDEFTQGAUTSOMQLNKILKYVDIGKN 384
QY 362 EGATLITGGERLGSKGYFIKPTVFGVDKEDMPVKEEIPGPVVITKPKSADEVINMAN 421
Db 385 EGATLITGGERLGSKGYFIKPTVFGVDKEDMPVKEEIPGPVVITKPKSADEVINMAN 444
QY 422 SEYGLAAGIHTSNINTALKVADRVNAGTWTWINTYNDPFHHAHPFGGFNASGLGREGMSVDA 481
Db 445 SEYGLAAGIHTSNINTALKVADRVNAGTWTWINTYNDPFHHAHPFGGFNASGLGREGMSVDA 504

QY 482 ONYLOVKAVRAKLD 495
Db 505 ONYLOVKAVRAKLD 518

RESULT 3
US-10-369-493-1749
: Sequence 1749, Application US/10369493
: GENERAL INFORMATION:
: APPLICANT: Cao, Yongwei
: APPLICANT: Hinkle, Gregory J.
: APPLICANT: Slater, Steven C.
: APPLICANT: Goldman, Barry S.
: APPLICANT: Chen, Xianfeng
: TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
: FILE REFERENCE: 38-10(52052)B
: CURRENT APPLICATION NUMBER: US/10/369,493
: PRIOR FILING DATE: 2003-02-28
: PRIOR APPLICATION NUMBER: US 60/360,039
: NUMBER OF SEQ ID NOS: 47374
: SEQ ID NO 1749
: LENGTH: 520
: TYPE: PRT
: ORGANISM: Saccharomyces cerevisiae
US-10-369-493-1749

Query Match 67.8% Score 1728.5; DB 6; Length 520;
Best Local Similarity 64.8%; Pred. No. 5, 1e 141;
Matches 320; Conservative 81; Mismatches 92; Indels 1; Gaps 1;

QY 2 SHLPMTVPKLPNGLEVEQPTGLFINNKEVPVSKQNKTEFVINSPEEFIECHIEGREDV 61
Db 24 SQAPLAVPTLDPNGFTYEQPTGLFINNKEVPVSKQNKTEFVINSPEEFIECHIEGREDV 84
QY 62 BEAVQAADRAFNSGWSNGIDPDRGKALYPLAELIEQDKDVIASITLDNGKATSSRGD 121
Db 84 BEAVQAADRAFNSGWSNGIDPDRGKALYPLAELIEQDKDVIASITLDNGKATSSRGD 142
QY 122 VDLVINYLKSSAGFADKIDGRMIDTGRTHFSYTKRQPLGVCGQIIPWNPFLMLMAWKIAP 181
Db 143 VALVSKYLRSGGWAIRKIVGNDIGKNIHIFYSIKPELAVAGQIIPWNPFLMLMAWKIAP 202
QY 182 ALVTGNTVVLKTAESTPLSALYVSKYIPQAGIPGVINIVSGFGKIVVEFATNHPKIKV 241
Db 203 ALAIGNTVVLKTAESTPLSALYVSKYIPQAGIPGVINIVSGFGKIVVEFATNHPKIKV 262
QY 242 AFTGSTATGRHIYQSAAGLKKVTLLELGGKSNIVFADAEKLVAVONTILGIYNSGEVC 301
Db 263 AFTGSTATGRHIYQSAAGLKKVTLLELGGKSNIVFADAEKLVAVONTILGIYNSGEVC 322
QY 302 CAGSRVYVEESIVDKFIEEFKAASESIKVGDPDFDEFTQGAUTSOMQLNKILKYVDIGKN 361
Db 323 CAGSRVYVEESIVDKFIEEFKAASESIKVGDPDFDEFTQGAUTSOMQLNKILKYVDIGKN 382
QY 362 EGATLITGGERLGSKGYFIKPTVFGVDKEDMPVKEEIPGPVVITKPKSADEVINMAN 421
Db 383 EGATLITGGERLGSKGYFIKPTVFGVDKEDMPVKEEIPGPVVITKPKSADEVINMAN 442
QY 422 SEYGLAAGIHTSNINTALKVADRVNAGTWTWINTYNDPFHHAHPFGGFNASGLGREGMSVDA 481
Db 443 SEYGLAAGIHTSNINTALKVADRVNAGTWTWINTYNDPFHHAHPFGGFNASGLGREGMSVDA 502
QY 482 ONYLOVKAVRAKLD 495
Db 504 SNTOTKRSVRJAD 516

RESULT 4
US-10-369-493-22495
: Sequence 22495, Application US/10369493

```

; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 22495
; LENGTH: 500
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-22495

Query Match 56.9%; Score 1452; DB 6; Length 500;
Best Local Similarity 54.8%; Pred. No. 1.2e-108;
Matches 272, Conservative 92, Mismatches 128, Indels 4, Gaps 3,

Qy 3 HLPMTVPIK--LPNGLEYQPTGLFINNKFPVSKONKTFEINPSTEEBEEICHIEGREDD 60
Dy 5 HPTAPPVKLTLPNGLYEQPTGLFINNKFPKKAQOGKTYPVDEPSTENTVCEVSATTED 64

Qy 61 VEVAAQAAPAFNSGWSNGIDPDGKALYRIARLTQADKIVTASTETLDNGKAISSSPG 120
Dy 65 VEVAIEADPAFHOTETAWQDPPEKPLLSKLADELESQIDLVSSIEALDNGKTLALRG 124

Qy 121 DVDLVNLYLKSSAGFADKIDGRMIDTGRTHFSYTRKPLGVCQIIIPNFPPLLMWANKIA 180
Dy 125 DVTLINCLURDAAYADKYNRTINTGDYMNFTLEPIGVCQIIIPNFPPLLMWANKIA 184

Qy 181 PALVTGNTVYLKTAESTPLSALVSKYIPQAGIPPGVINIVSGFKIVVEAITNHPKIKK 240
Dy 185 PALAMGNVCLKPAVTPLNALYFASLCKKVGIPAGVNVIPGRTVGAALTNDPRIR 244

Qy 241 VAFGTSTATGRHI-YOSAAAGLKVKLTLELGGKSPNIVFADAEKKAQONIIILGYNSGE 299
Dy 245 LAFGTSTGVKSVAVDSSSENLKITLELGGKSAHLVFDGANIKKLEPNLVNGIFKNAGQ 304

Qy 300 VCCAGSPVVFYFESYKPTFEKKAASES-IKVGPEPDESTFGAQTQSOMQLNKILKYVDI 358
Dy 305 ICSSGRIYVOEGLYDELLAAFAKAYLETEIKVGNFPDKANTFGAQTNRQQPDTIMNYIDI 364

Qy 359 GKNEGATLITGERLGSKGYFIKPTVFGDVKEDMRIVKEEIPGPVVTVTKFSADEVINM 418
Dy 365 GKKEGAKILTGGKVGKGYFIRPTVFYDVNEDMRIVKEEIPGPVVTVAKEKTEEGVEM 424

Qy 419 ANUSEYGLAAGIHTSNINIALKVAUKVNAQTWIMNYNLPFHIAVPPFGFNASGLGREMSV 478
Dy 425 ANSEFLGSGLETSLTGLKVAKMLKAGTWIMTYNDVSRVPPFGVKGUSGYGREMGE 484

Qy 479 DALQNYLOKVAVRKL 494
Dy 485 EVYHAYTEVKAVRIKL 500

```

RESULT 5

```

US-10-369-493-3846
; Sequence 3846, Application US/10369493
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B

```

RESULT 6

```

US-10-219-051B-6969
; Sequence 6969, Application US/10219051B
; GENERAL INFORMATION:
; APPLICANT: The General Hospital Corporation doing business as Massachusetts General
; APPLICANT: Hospital / Bayer AG
; TITLE OF INVENTION: Nucleotide sequences involved in pain
; FILE REFERENCE: Lea 35693 Foreign Countries
; CURRENT APPLICATION NUMBER: US/10/219,051B
; CURRENT FILING DATE: 2003-05-09
; PRIOR APPLICATION NUMBER: US 60/312,147
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: US 60/346,382
; PRIOR FILING DATE: 2001-11-01
; PRIOR APPLICATION NUMBER: US 60/333,347
; PRIOR FILING DATE: 2001-11-26
; NUMBER OF SEQ ID NOS: 14715

```

```

; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3846
; LENGTH: 498
; TYPE: PRT
; ORGANISM: Neurospora crassa
; NAME/KEY: unsure
; LOCATION: (1)-(498)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-3846

Query Match 56.2%; Score 1432.5; DB 6; Length 498;
Best Local Similarity 54.1%; Pred. No. 4.4e-107;
Matches 268, Conservative 88, Mismatches 114, Indels 25, Gaps 3,

Qy 25 FINNFVPSKUNKTFEINPSTEEBEEICHIEGREHVEEAVGAAGAFNSN3SWNGTDPID 84
Dy 1 FINNEFVGVKKTFEINPATEEIVCSVHEAEKDVDAVAAAKAF-E3VWKDVTPQQ 59

Qy 85 RGKALYRLAELEQDKDVIASTETLDNGKAISSSRGVDLVLYIKSSAGFADKIDGRMI 144
Dy 60 RGIYLLKLLADLLEKLLDLAAVESLDNGKSIIMARGDVGAVVGTIRYYGWDKIEGKTI 119

Qy 145 DTGRTHFSYTRKPLGVCQIIIPNFPPLLMWANKIAPALVTGNTVVLKTAESTPLSALV 204
Dy 120 DISPSDFHYTROEPLGVCQIIIPNFPPLLMWANKYGPALATGNTIVMKTABQTPLSALV 179

Qy 205 SKYIPQAGIPPGVINIVSGFKIVVEAITNHPKIKKVAFTGSTATGRHIYOSAA-AGLKK 263
Dy 180 AGFVKAGPPGVNLIISGFGRIAGAAASHMDINIKVAFSTGTVGRLMKAAAESNLKK 239

Qy 264 VTLELGGKSPNIVFADAEKKAQONIIILGYNSGEVCCAGSPVYVEEISYDKTFEEFKA 323
Dy 240 VTLELGGKSPNIFENDADIDQIDAVNFGIVNHCQTCCAGSRVYVQEGIYDKFVAAPKQ 299

Qy 324 ASESIVKGPDPFDESIFQGAQTQSOMQLNKILKYVDIGKNEGATLITGGERLGSKGYFIKPT 383
Dy 300 RAQCNKVPVPHOETFGPQVSQLAYOPIMRYIKAKKEEATVETGRRHDKGYFQPT 359

Qy 384 VFGDVKEDMRIVKEEIPGPVVTVTKFSADEVINNMANSSEYGLAAGIHTSNINTALKVAD 443
Dy 360 IFINVPHLMLKKEELFSPCAVAKFSTFEFVILGNLSNYGLAAAVHTKDLTAIRVN 419

Qy 444 RVNAGTVTWT-----YNDPHHAVPPGFGFNASGLGREMSVDA 480
Dy 420 HLRAGTVVWNTCKYTCFVLEGVTFHRANONLFQOTNALHQLPFGYKESGIGRELGEAA 479

Qy 481 LQNYLOKVAVRKL 495
Dy 480 LANYTCKSVAKLN 494

```

```

; SOFTWARE: Perl script
; SEQ ID NO 6969
; LENGTH: 519
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: SWISS-Prot / P11884
; DATABASE ENTRY DATE: 2002-06-15
US-10-219-051B-6969

```

```

Query Match      50.0%, Score 1274.5, DB 6, Length 519,
Best Local Similarity 50.8%, Pred. No. 2 8e 94,
Matches 252, Conservative 85, Mismatches 146, Indels 13, Gaps 6;

QY 7 TVPIKLPNGLEYEQP-----TGLFINNKFPVPSKONKTFFVINPSTEEETICHIEYEGREDDVE 62
DQ 24 TSAVPAPN-----QOPEVFCNQIFINNEWHDAVSKKTFPTVNPSTGEVICOVAEGNKEDVD 79
QY 63 EAVQAADRAFNSGS--WNGIDPIDRGKALYLAELIADODKDVIASTETLDNGKA-ISSSKG 120
DQ 80 KAVKAQAQAFQLGSPWRMDASDRGLLYRLADLIERDRTYLAALETLDNGKPYVIVSLV 139
QY 121 DVDLVINYLKSSAGFADKIDGRMIDTGRTHFSYTKRQPLGVCQIIPWNPFLMMAWKIA 180
DQ 140 DLDVVLKCLRYAYAGWADKYHGKTIPIIDGDFFSYTRHEPVGVCQIIPWNPFLMQAWKLG 199
QY 181 PALVTGNTVYLKTAESTPLSALVSKYIPQAGIPPGVINIVSGFGKIIVVEAITNHPKTK 240
DQ 200 PALATGNVVMKVAEQPTLALYVANLIKEAGFPVGVNIVPGCPGTAGAAIASHEDVDK 259
QY 241 VFTGSTATGRHLYQSAA--AGLKVYTLGLGKSNIVFADAEKKAQVNIILGIYYNSG 298
DQ 260 VFTGSTEVG-HLIQVAAGSSNLKRVTLGLGKSNIVFADAEKKAQVNIILGIYYNSG 318
QY 299 EVCCAGSRVYVESIYDKFTFEERKAASESIKVGDDPDESTFGCAQTSQMLNKLKYYDI 358
DQ 319 QCCAGSRTEVQEDVDYDEFVERSAKSKVGNIPDSRTEQGPQVDETFQKKILGYIKS 378
QY 359 GKNEGATLITGGERLGSKGYFIKPTVFGDVKEDMRIVKEEIPGVVTVTKFSADDEVINM 418
DQ 379 GQEGAKLLCGGGAADAAAFYIQPTVFGDVKEDMTIAKEEIPGVVQILKFTIEEVGR 438
QY 419 ANDSEYGLAAGIHTSNITALKVADRVNACTVWINTYNDPHAVPFGGFGNAGSLGRMSV 478
DQ 439 ANSKYGLAAAVFTKLDKANYLSQALQAGTVMINCYDVFQAQSPFGGKMSGSGRELGE 498
QY 479 DALONYLQVKAARAKL 494
DQ 499 YGLQAYTEKVTIVKV 514

```

```

RESULT 7
US-10-219-051B-6973
; Sequence 6973, Application us/10219051B
; GENERAL INFORMATION:
; APPLICANT: The General Hospital Corporation doing business as Massachusetts General
; APPLICANT: Hospital / Bayer AG
; TITLE OF INVENTION: Nucleotide sequences involved in pain
; FILE REFERENCE: Lea 35693 Foreign Countries
; CURRENT APPLICATION NUMBER: US/10/219,051B
; CURRENT FILING DATE: 2003-05-09
; PRIOR APPLICATION NUMBER: US 60/312,147
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: US 60/346,382
; PRIOR FILING DATE: 2001-11-01
; PRIOR APPLICATION NUMBER: US 60/333,347
; PRIOR FILING DATE: 2001-11-26
; NUMBER OF SEQ ID NOS: 14715
; SOFTWARE: Perl script
; SEQ ID NO 6973
; LENGTH: 519
; TYPE: PRT
; ORGANISM: Rattus norvegicus

```

```

; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: SWISS-Prot / P11884
; DATABASE ENTRY DATE: 2002-06-15
US-10-219-051B-6973

```

```

Query Match      50.0%, Score 1274.5, DB 6, Length 519,
Best Local Similarity 50.8%, Pred. No. 2 8e 94,
Matches 252, Conservative 85, Mismatches 146, Indels 13, Gaps 6;

QY 7 TVPIKLPNGLEYEQP-----TGLFINNKFPVPSKONKTFFVINPSTEEETICHIEYEGREDDVE 62
DQ 24 TSAVPAPN-----QOPEVFCNQIFINNEWHDAVSKKTFPTVNPSTGEVICOVAEGNKEDVD 79
QY 63 EAVQAADRAFNSGS--WNGIDPIDRGKALYLAELIADODKDVIASTETLDNGKA-ISSSKG 120
DQ 80 KAVKAQAQAFQLGSPWRMDASDRGLLYRLADLIERDRTYLAALETLDNGKPYVIVSLV 139
QY 121 DVDLVINYLKSSAGFADKIDGRMIDTGRTHFSYTKRQPLGVCQIIPWNPFLMMAWKIA 180
DQ 140 DLDVVLKCLRYAYAGWADKYHGKTIPIIDGDFFSYTRHEPVGVCQIIPWNPFLMQAWKLG 199
QY 181 PALVTGNTVYLKTAESTPLSALVSKYIPQAGIPPGVINIVSGFGKIIVVEAITNHPKTK 240
DQ 200 PALATGNVVMKVAEQPTLALYVANLIKEAGFPVGVNIVPGCPGTAGAAIASHEDVDK 259
QY 241 VFTGSTATGRHLYQSAA--AGLKVYTLGLGKSNIVFADAEKKAQVNIILGIYYNSG 298
DQ 260 VFTGSTEVG-HLIQVAAGSSNLKRVTLGLGKSNIVFADAEKKAQVNIILGIYYNSG 318
QY 299 EVCCAGSRVYVESIYDKFTFEERKAASESIKVGDDPDESTFGCAQTSQMLNKLKYYDI 358
DQ 319 QCCAGSRTEVQEDVDYDEFVERSAKSKVGNIPDSRTEQGPQVDETFQKKILGYIKS 378
QY 359 GKNEGATLITGGERLGSKGYFIKPTVFGDVKEDMRIVKEEIPGVVTVTKFSADDEVINM 418
DQ 379 GQEGAKLLCGGGAADAAAFYIQPTVFGDVKEDMTIAKEEIPGVVQILKFTIEEVGR 438
QY 419 ANDSEYGLAAGIHTSNITALKVADRVNACTVWINTYNDPHAVPFGGFGNAGSLGRMSV 478
DQ 439 ANSKYGLAAAVFTKLDKANYLSQALQAGTVMINCYDVFQAQSPFGGKMSGSGRELGE 498
QY 479 DALONYLQVKAARAKL 494
DQ 499 YGLQAYTEKVTIVKV 514

```

```

RESULT 8
US-10-369-493-12695
; Sequence 12695, Application US/10369493
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12695
; LENGTH: 532
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)-(532)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-12695

```



```

Db 78 KAVKARAAPOLASPMWPMDASHRGPIINPLADLIERDPTYLAALLETLDNGKPPYVSYLV 137
QY 121 DVDIVINYLKSSAGFADKIDGRMIDTCRTHFSYTKRQPLGVCQGIIPWNPFLMLMANKIA 180
Db 138 DLDWVLCRLRYAGWADKYHGKTIPIIDGDFSYRHEPVGVCQGIIPWNPFLMLMANKLG 197
QY 181 PALVTGNTVVLKTAESTPLSALYSKYIIPQAGIIPGVINIVSGFGKLVVEALTNHPKIKK 240
Db 198 PALATGNVVMKVAEQTPLTALYANLKEAGFPFGVNVIVPGFPTAGAAIASHEDVDK 257
QY 241 VAFTGSTATGRHHYQSA-AAGLKKVTLGLGKSPNIVFADAEKKAQVNIILGIYNSGE 299
Db 258 VAFTGSTEIGRVQVAAGSSNLKRVTLGLGKSPNIMSDADMWAVEQAHFALFENQGG 317
QY 300 VCCAGSVVYVESIYDKFIEEFKAASESIKVGDPDEDFQCAQTSQMLNKLKAYVDIG 359
Db 318 CCCAGSRTFVQEDYDEFVRSVARAKSRVGNPFDSKTEQGPQVDETFQFKILGIYNTG 377
QY 360 KNEGATLITGGERLGSKGFIKPTVFGVYKEDMRIVKEEIFGPVVTVTKFSADDEVINNA 419
Db 378 KQEGAKLGGGLAADKGYFIQPTVFGVQDGMTIAKEEIFGPVVMILKFKTIEEVVGR 437
QY 420 NDSEYGLAAGIHTSNINTALKVADRVNAGTVMWINTYNDHHAVPFGGPNASGLGHEMSVD 479
Db 438 NNSTYGLAAAVFTKDLKDKANYLSQALQAGTVMWNYDVFAGUSPFGYKMSGSGHELG 497
QY 480 ALQNYLOVKAVRAKL 494
Db 498 GLQAYTEVKTVTKV 512

RESULT 11
US-09-949-016-10653
; Sequence 10653, Application US/09949016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10653
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10653

```

```

Query Match 49.2%; Score 1253.5; DB 5; Length 520;
Best Local Similarity 50.3%; Pred. No. 1.4e-92;
Matches 249; Conservative 86; Mismatches 149; Indels 11; Gaps 5;

QY 7 TVPIKLPNGIYEQPTGLFINNKFPVSKONKTFEIVNPFSTEIEICHIEGREDDE 62
Db 25 TQAVPAPN---QOPEVFCNQIFINNEHDAVSRTFTFVPSTGEVICQVAEGDKEDVD 80
QY 63 EAVQADAPAFSNGS-WNGIDPTDRKALYRLAELEQDKDVIASTETLDNGKA-ISSRG 120
Db 81 KAVKARAAPOLGSPWRRMDASHRGLLNRLADLIERDRTYLAALLETLDNGKPPYVSYLV 140
QY 121 DVDIVINYLKSSAGFADKIDGRMIDTCRTHFSYTKRQPLGVCQGIIPWNPFLMLMANKIA 180
Db 141 DLDWVLCRLRYAGWADKYHGKTIPIIDGDFSYRHEPVGVCQGIIPWNPFLMLMANKLG 200
QY 181 PALVTGNTVVLKTAESTPLSALYSKYIIPQAGIIPGVINIVSGFGKLVVEALTNHPKIKK 240

```

```

Db 201 PALATGNVVMKVAEQTPLTALYANLKEAGFPFGVNVIVPGFPTAGAAIASHEDVDK 260
QY 241 VAFTGSTATGRHHYQSA-AAGLKKVTLGLGKSPNIVFADAEKKAQVNIILGIYNSGE 299
Db 261 VAFTGSTEIGRVQVAAGSSNLKRVTLGLGKSPNIMSDADMWAVEQAHFALFENQGG 320
QY 300 VCCAGSVVYVESIYDKFIEEFKAASESIKVGDPDEDFQCAQTSQMLNKLKAYVDIG 359
Db 321 CCCAGSRTFVQEDYDEFVRSVARAKSRVGNPFDSKTEQGPQVDETFQFKILGIYNTG 380
QY 360 KNEGATLITGGERLGSKGFIKPTVFGVYKEDMRIVKEEIFGPVVTVTKFSADDEVINNA 419
Db 381 KQEGAKLGGGLAADKGYFIQPTVFGVQDGMTIAKEEIFGPVVMILKFKTIEEVVGR 440
QY 420 NDSEYGLAAGIHTSNINTALKVADRVNAGTVMWINTYNDHHAVPFGGPNASGLGHEMSVD 479
Db 441 NNSTYGLAAAVFTKDLKDKANYLSQALQAGTVMWNYDVFAGUSPFGYKMSGSGHELG 500
QY 480 ALQNYLOVKAVRAKL 494
Db 501 GLQAYTEVKTVTKV 515

RESULT 12
US-10-369-493-23162
; Sequence 23162, Application US/10469493
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongqiao
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)R
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 23162
; LENGTH: 495
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US 10 369 493-23162

Query Match 49.1%; Score 1252.5; DB 6; Length 495;
Best Local Similarity 49.0%; Pred. No. 1.5e-92;
Matches 243; Conservative 92; Mismatches 158; Indels 4; Gaps 2;

QY 1 MSHLPMTVPPIKLPNGIYEQPTGLFINNKFPVSKONKTFEIVNPFSTEIEICHIEGRED 60
Db 1 MSSLTMQVTKRLETFQGTGKK-LYIDGKFVVSASCATFTDTPNATGCTELMTEYEAQAD 58
QY 61 VEEAVQADAPAFSNGS-WNGIDPTDRKALYRLAELEQDKDVIASTETLDNGKA-ISSR 119
Db 59 VDKAVKAARKAPDQGEWRTMSPASRSRLMYKLADLMEEHKTLELAQLETLDNGKINETN 118
QY 120 GVDVLVINYLKSSAGFADKIDGRMIDTCRTHFSYTKRQPLGVCQGIIPWNPFLMLMANKI 179
Db 119 GDPIALEHMYRYAGWCWKITQUTIPVSCATFNTHEPVGVCQGIIPWNPFLMLMANKM 178
QY 180 APALVTGNTVVLKTAESTPLSALYSKYIIPQAGIIPGVINIVSGFGKLVVEALTNHPKIK 209
Db 179 GAALATGCTEVLKPAQGPPLSALYLAELIPGAGFAGVNIIPGFGFAGALTNHGAVD 238
QY 240 KVAFTGSTATGRHHYQSAAGLKKVTLGLGKSPNIVFADAEKKAQVNIILGIYNSGE 299
Db 236 KIATFQSTEDPKINSTAAKSTKRVTLGLGKSPNIMSDADMWAVEQAHFALFENQGG 298
QY 300 VCCAGSVVYVESIYDKFIEEFKAASESIKVGDPDEDFQCAQTSQMLNKLKAYVDIG 359
Db 299 VVACASPVTHKDTYFVVYFMASVAESLKEAVLJIKHIGLHVLKHLFELVLYAGG 398

```

```

QY 360 KNEGATLITGGERLGSCKYFTKPKVPLGWLKEDMPVKEEIPGVVIVTKFKSADEVINMA 419
Db 359 KDEGAKAVTGGSCPFAGYFVAPTVFANVEDEMTIAKEEIFGVLTAIPYETVDEVIERA 418
QY 420 NDEYGLAAGTHTSNTALKVADRNVAGTWTWINTYNDFIHIAVPEGCFNASGLGKREMSVD 479
Db 419 NHSEYGLAAGLTENVKQAHYIADRLQAGTIVVNCYNVFDAAAPFGYKOSGLGREGMSY 478
QY 480 ALONYLOVKAVRAKID 495
Db 479 ALDNYTEVKSVMWNL 494

RESULT 13
US-10-369-493-3555
; Sequence 3555, Application US/10369493
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Harry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10/520521B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 697,460,049
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3555
; LENGTH: 516
; TYPE: PRT
; ORGANISM: Neurospora crassa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(516)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-3555

```

```

Query Match 48.8%; Score 1243.5; DB 6; Length 516;
Best Local Similarity 47.3%; Pred. No. 8.7e-92;
Matches 342; Conservative 92; Mismatches 155; Indels 23; Gaps 4;

QY 6 MTPVILKPLNGLEYEPTGLFINNKFPVSKQNKTFEVIN-----PST 46
Db 5 MEVELTAPNGKKWQPLGLFINNEFVKSANEQKLISINPTVYVSNPISLTXLRVLTIPST 64
QY 47 EEECHIVYEGREDDVEEAQVADRAFNGSWGNGIDPIDRGKALYRIAELEQDKWIASI 106
Db 65 EEEICSVYAATAEDVDAAVSAARAFRHSWKSLSGTGERGALMKPLADLVAENAEILATI 124
QY 107 ETLDNCKALSSRGD-VDLVTNYLKSSAGFAUKIPDEMIDTGHSTHSYTKPQPLGVCQOI 165
Db 125 ECLDNCKPQYQALNENPEVITNVLRYAGYADKNFGQVIDGPAKFAYTVKEPLGVCQOI 184
QY 166 IPWNPPLLMWAKTAPALVTGNTVLTABSTPLSALYVSKYIPQAGCIPPGVINIVSGFG 225
Db 185 IPWNPPLDMAAKLGPALCGNTVVLKAEQIPLSVLYLAKLKEAGFPVSVNLINGH 244
QY 226 KIVWEATINHPKIKKVAFTGTATGRHIIYOSAAAGLKKVTLGCGKSPNIVFADAEKKA 285
Db 245 RPAGAAIVQHPQVTKIAPTGTGTEKIMKNMASYTMKNITLETGKSPITVEFDADLELA 304
QY 286 VQNIILGIYNSGWCAGSVVYVEESLYDKFTEEPKAASESISKV-GDPDESTFOGAOT 344
Db 305 ATWSHIGIMNSGOQICATSRILVBEKIYDEFEKFAKVOEVSVLGDPEESTFRGPQV 364
QY 345 SUMQUNKILKYVDIGKNNGATLITIGGEPL--GSGNGYFKPTVFSLVKEDMPVKEEIPGP 402
Db 365 TKAQYERVLGVINYSKKEGATVMGQTEPAQNSKSGFFVAPIVFTNVKPTMKIPPEKIPGP 424

```

```

QY 433 VVTVTKFKSALEVINMANUSFYGLAASTHTSNTINTALKVADRYNAATVWINTYNDHFHAV 462
Db 425 QVALITFKTEBEALTLANDSMYSLGAALPTFKTTRAFHPVAPRIEAGMWWVNSNSDSDRF 484
QY 463 PFGGPNASGLGKREMSVDALQNYLQVKAVRAKL 494
Db 485 PFGSVKUSNIGFELGEACLAPYCNVKSITHVNL 516

RESULT 14
US-60-453-135-14871
; Sequence 14871, Application US/60453135
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: IAKOUBOVA, Olga
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: MYOARIAL INFACTION, METHODS OF TREATMENT AND USES THEREOF
; FILE REFERENCE: CU001456
; CURRENT APPLICATION NUMBER: US/60/453,135
; CURRENT FILING DATE: 2003-03-10
; NUMBER OF SEQ ID NOS: 82762
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14871
; LENGTH: 514
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-453-135-14871

Query Match 48.6%; Score 1239.5; DB 7; Length 514;
Best Local Similarity 50.4%; Pred. No. 1.8e-91;
Matches 246; Conservative 85; Mismatches 146; Indels 11; Gaps 5;

QY 7 TVPIKLPLNGLEYEQP----TGLFINNKFPVSKQNKTFEVINPSTEBEICHYEGREDDVE 62
Db 22 TQAVFAFN-----QQFEVFCNQIFINNEHDAVSKRTFTVNPSTGEVICOVAGDREDVD 77
QY 63 EAVQAADPAFNGS-WNGIDPIDRGKALYRIAELEQDKWIASIETIUNGKA-ISSSPG 120
Db 78 KAVKAARAAAFOLGSPWPPMNASHPGLNLRLADLIERDRTYLALETLDNGKPYVISYL 137
QY 121 DVLIVNYLKSSAGFADKIDGPMIDTGHSTHSYTKPQPLGVCQOIIPWNPPLLMWAKIA 180
Db 138 DLDVLYKRLRYAGWADYHGKTIPIGIDFESTIRHEPVGVCQOIIPWNPPLLMQAWKLG 197
QY 181 PALVTGNTVLTABSTPLSALYVSKYIPQAGCIPPGVINIVSGFGKIVVEATINHPKIKK 240
Db 138 PALAIGNVWVWKKVAEQIPLTALYVANIKEAGFPVGVNIVPGFGTAGAAIASHEDVDK 257
QY 241 VAPTGTATGRHIIYOSA-AAELKKVTLGCGKSPNIVFADAEKKAQVNIITIGIYVNSGP 294
Db 258 VAPTGTSTEDGVLQVAAVSSNLKPKVTLEDGKSNILMSALMLWAVEQAHPALFENQJ 317
QY 306 VQCAKSVYVVEESLYDKFIEEFKAASESISKVGGEPDESIFGAGCTSQMLNKLILKYVDIG 359
Db 318 QCAKASPTVVEEDLYDEFPVHSAVAPAKSVVGNPFUSKIFGSPQVDETQFKKILGVTNG 377
QY 360 KNEGATLITGGERLGSCKYFTKPKVPLGWLKEDMPVKEEIPGVVIVTKFKSADEVINMA 419
Db 378 KQEGAKTITGGGTAAGAGCYEIPQTVFSLVQSGMIAKAFIEGDPVMATLKEETLEEVGRA 437
QY 420 NDEYGLAAGTHTSNTALKVADRNVAGTWTWINTYNDFIHIAVPEGCFNASGLGKREMSVD 479
Db 438 NNSYGLAAAVYETKRIKANYISAGAGATVWVWNYIAPVTAQSEETCYKMSGSGREIGPY 497
QY 480 ALONYLOV 487
Db 498 GLQAYTEV 505

RESULT 15
US-60-453-050-14871
; Sequence 14871, Application US/60453050
; GENERAL INFORMATION:

```

```
; APPLICANT: CARGILL, Michele
; APPLICANT: LUKE, May
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE OF INVENTION: SIENOSIS, METHODS OF DETECTION AND USES THEREOF
; CURRENT APPLICATION NUMBER: CL001457
; CURRENT FILING DATE: 2003-03-10
; NUMBER OF SEQ ID NOS: 82762
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14871
; LENGTH: 514
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-453-050-14871

Query Match.      48.6%; Score 1239.5; DB 7; Length 514;
Best Local Similarity 50.4%; Pred. No. 1.8e-91;
Matches 246; Conservative 85; Mismatches 146; Indels 11; Gaps 5;

QY   7  TVPIKLPNGLEYRPP---TGLFINNKFPVPSKQNKTFEIVNPSTREEICHIVEGREDDVE 62
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
22  TOAVPAEN ...QQPEVFCNCFINNEWHDAVSRKTFEFTVNESTGEVICQVAEGDKREDVD 77
QY   63  EAVOAAADRAFNSGS-WNGIDIPIDRGKALYRLAELIFODKDVIASTETLDNGKA-ISSSRG 120
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
78  KAVKAARAAPQLGSPWRMRMDASHRGRLNKLADLIERDRTYLAALETLDNGKPYVISYLV 137
QY   121 DVDLVINYLKSSAGPADKIDGRMIDGRTHFSYTKRQPLGVCQGIIPNFFPLLMKAWKIA 180
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
138  DLDVLKLCRYAGWADKYHGKTIPIIDGFFSYTRHEPVGCGQIIPNFFPLLMQAWKLG 197
QY   181 PALVTGNVTVLKTAESTPLSALYVSKYIPQAGIPPGVINIVSGFGK1VVEAITNHPKIKK 240
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
198  PALATCNVVMKVAEQTPLTALYVNLKEAGFPQGVNIVPGFPTAGAAIASHEDVDK 257
QY   241 VAFGTGTATGRHIYQSA-AAGLKVYTLLELGGKSPNIVFADAEKKAQVNIILGIYYNSGE 299
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
258  VAFGTGSTEIGRVTOVAAAGSSNLKRVTLLELGGKSPN1MSDADMDWAVEQAHFALFFNOG 317
QY   300 VCCAGSRVYVEESIYDKFTEEFKAASESTKVGDPDESTFOGAGTSSOMLNKILKYVDIG 359
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
318  CCCAGSKTFVORDIYDEFVRSVARAKSRVGNPFDKIEQQPVQDETQFKKILGYINTG 377
QY   360 KNEGATLITGGHLEKSGGYFIKPTVFGDVKEDMPIVKEEIEGPPVVVTIKPKSADEVINMA 419
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
378  KQEGAKLLCGGGLAADRGVFIQPTVFGDVQDCMTIAKEEIEGPPVMQILKPKTIEEVVGRA 437
QY   420 NDSEYGLAAGIHTSNINTALKVADRVNAGTVMINTYNDFHHAVPFGGENASGLGKREMSVD 479
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
438  NNSTYGLAAVFTKDLKANYLSQALQAGTVWVNCYDVFAGOSPFGGYKMSGSGRELGEY 497
QY   480 ALQNYIQV 487
Db   : : : : :
498  GLQAYTEV 505
```

Search completed: June 24, 2003, 10:35:19
Job time : 54.5876 secs

GenCore version 5.1.6
Copyright (c) 1991 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:10:05 : Search time 18.9426 Seconds
(without alignments)
2512.144 Million cell updates/sec

Title: us-09-830-751-2
Perfect score: 2550
Sequence: 1 MSLLPMTVPKLPNCLEVEQ MSVDALGNLYGVKAVPAKLD 495

Scoring table: BLOSUM62
Gapop (6 0) , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR_73: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	Query Match %	Length	DR	ID	Description
1	2538	99.5	519	1	S67286	probable aldehyde
2	1728.5	67.8	520	1	S05076	probable aldehyde
3	1452	56.9	500	1	S69929	probable aldehyde
4	1412.5	55.4	497	1	A29055	aldehyde dehydroge
5	1399	54.9	496	2	S43114	aldehyde dehydroge
6	1362.5	53.4	495	2	S43108	aldehyde dehydroge
7	1362	53.4	503	2	T39216	aldehyde dehydroge
8	1275.5	50.0	519	1	I48966	aldehyde dehydroge
9	1274.5	50.0	519	1	S03564	aldehyde dehydroge
10	1253.5	49.2	517	1	DEHUE2	aldehyde dehydroge
11	1252.5	49.1	445	2	H63614	aldehyde dehydroge
12	1252.5	49.1	517	1	A40872	aldehyde dehydroge
13	1240.5	48.6	520	1	S00040	aldehyde dehydroge
14	1236	48.5	544	2	D88449	protein F54D8.3 (i
15	1232.5	48.3	549	2	T03983	rif2 nuclear restor
16	1222.5	47.9	499	2	S74224	aldehyde dehydroge
17	1220.5	47.9	500	1	S00364	aldehyde dehydroge
18	1218.5	47.8	512	1	A35684	aldehyde dehydroge
19	1216.5	47.7	542	2	T02301	aldehyde dehydroge
20	1212.5	47.5	509	1	S14639	aldehyde dehydroge
21	1205.5	47.3	498	2	CR3717	NADP-dependent ald
22	1201	47.1	538	2	T06683	aldehyde dehydroge
23	1192.5	46.8	497	2	T04924	aldehyde dehydroge
24	1189	46.6	519	2	C86372	hypothetical prote
25	1182.5	46.4	501	1	DEHFE1	aldehyde dehydroge
26	1181.5	46.3	501	2	S14752	aldehyde dehydroge
27	1179.5	46.3	501	2	J05553	aldehyde dehydroge
28	1176.5	46.1	500	1	S02302	aldehyde dehydroge
29	1172.5	46.0	496	2	T00077	probable aldehyde

30	1170.5	45.9	501	1	I61004	aldehyde dehydroge
31	1169.5	45.9	501	1	A32616	aldehyde dehydroge
32	1168.5	45.8	501	2	J04524	aldehyde dehydroge
33	1141	44.7	498	1	S43184	aldehyde dehydroge
34	1101.5	43.2	496	1	A46725	omega-crystallin -
35	1084	42.5	494	2	B95411	probable aldehyde
36	1081	42.4	506	1	S54615	aldehyde dehydroge
37	1057	41.5	506	1	S54527	aldehyde dehydroge
38	1034.5	40.6	490	2	A64639	glycine betaine al
39	1020	40.0	495	2	C89778	hypothetical prote
40	1006	39.5	511	1	S41308	aldehyde dehydroge
41	1002.5	39.3	500	2	AH4544	betaine aldehyde d
42	976	39.1	496	2	C87638	aldehyde dehydroge
43	973	38.9	492	2	A60560	formyltetrahydrofo
44	992	38.9	495	2	H83136	probable aldehyde
45	976.5	38.3	507	2	AG3187	aldehyde dehydroge

ALIGNMENTS

RESULT 1

S67286

probable aldehyde dehydrogenase (NAD) (EC 1.2.1.3) YOR374W - yeast (Saccharomyces cer
N: Alternate names: hypothetical protein U6730
C: Species: Saccharomyces cerevisiae
C: Date: 12-Jul-1996 #Sequence_revision 41-Jan-1997 #text_change 04-Jun-2002
C: Accession: S67286

R: Delius, H.; Hebling, U.; Hofmann, B.
submitted to the Protein Sequence Database, July 1996
A: Reference number: S67261

A: Accession: S67286

A: Molecule type: DNA

A: Residues: 1-519

A: Cross references: EMBL:Z75282; PIDN CAA99705.1; GSPDR GND00015; MIPS:YOR374W

A: Experimental source: strain S288C

C: Genetics:

A: Gene: SGD:ALD7; MIPS:YOR374W

A: Cross references: SGD:S0005901; MIPS:YOR374W

A: Map position: 15R

C: Function:

A: Description: catalyzes oxidation of an aldehyde to an acid using NAD+ and water
A: Note: enzymes with this activity are involved in diverse metabolic pathways in vari
C: Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
C: Keywords: NAD, oxidoreductase
E: 81-344/Domain; aldehyde dehydrogenase homology
F: 290,324/Active site; Glu, Cys #status predicted

Query Match 99.5%; Score 2538; DB 1; Length 519;
Best Local Similarity 99.8%; Pred. No. 4.9e-159;
Matches 493, Conservative 0, Mismatches 1, Indels 0; Gaps 0;

Q7 2 SHLPMTVPKLPNCLEVEQPTGLFNNKVEFSKNGKTFEVINISTEEICHVCEHDEV 61

DB 25 SHLPMTVPKLPNCLEVEQPTGLFNNKVEFSKNGKTFEVINISTEEICHVCEHDEV 84

Q7 62 EFAVAAADRAFNSGNSWNGIDIDRGKALYLAELLEGKUVIASIELDNKALSSSGD 121

DB 85 EFAVAAADRAFNSGNSWNGIDIDRGKALYLAELLEGKUVIASIELDNKALSSSGD 144

Q7 122 VDLVINYLKSSAFAKIDGEMIDGTFESYKKEFGVAVGQIIPNFPPLMAMKAP 181

DB 145 VDLVINYLKSSAFAKIDGEMIDGTFESYKKEFGVAVGQIIPNFPPLMAMKAP 204

Q7 182 ALVTGNTVVKTAETISLVSLSYVSKYIQAGIFGVINIVSGFIVVLAIIHPRKEV 241

DB 245 ALVTGNTVVKTAETISLVSLSYVSKYIQAGIFGVINIVSGFIVVLAIIHPRKEV 264

Q7 242 AFTGSTATGRHYGSAAGLKEVTFEGGKSPNVFAACAFKAVCNHIGVYNSDEV 301

DB 265 AFTGSTATGRHYGSAAGLKEVTFEGGKSPNVFAACAFKAVCNHIGVYNSDEV 324

Q7 302 GARGPVYVESLYKRFEEFKAAKSEIKVGGHFEHFEFTEFQAQLQMOLRIKLKLYVLGKN 361

Db 325 CAGSRVTVESIVDKFEEFKASESIKVGDPDESTFGAOTSOMQNKILIKYVDIGKN 384
 QY 362 EGATLITGGRIASGKGYFKPTVFGDVQKEIMRIVKEEIPGVVTVTKFKSADEVINMAND 421
 Db 385 EGATLITGGRIASGKGYFKPTVFGDVQKEDMRIVKEEIPGVVTVTKFKSADEVINMAND 444
 QY 422 SEYGLAAGIHTSNINTALKVADRVNAGTVWINTYNDPHFHAVPFGGFGNAGSLGRENMSVDAL 481
 Db 445 SEYGLAAGIHTSNINTALKVADRVNAGTVWINTYNDPHFHAVPFGGFGNAGSLGRENMSVDAL 504
 QY 482 QNYLOVKAVRAKLD 495
 Db 505 QNYLOVKAVRAKLD 518

RESULT 2

S50576

probable aldehyde dehydrogenase (NAD) (EC 1.2.1.3) YER073w yeast (Saccharomyces cerevisiae)

C:Species: Saccharomyces cerevisiae

C>Date: 28 May 1993 #sequence_revision 31-Jan-1997 #text_change 03-Jun-2002

C:Accession: S50576

R:Blotrich, F.S.

submitted to the EMBL Data Library, December 1994

A:Description: The sequence of S. cerevisiae lambda clone 3612 and cosmid 9747.

A:Reference number: S50438

A:Accession: S50576

A:Molecule type: DNA

A:Residues: 1-520 <DIE>

A:Cross references: EMBL:U08814; NID:G60400; PDB:AAK64612 1; PDB:9G0410; GSPDB:GN0000

C:Experimental source: strain S288C (AB972)

C:Genetics:

A:Gene: SGD:ALD5; MIPS:YER073w

A:Cross references: SGD:S0000875

A:Map position: 5P

C:Function:

A:Description: catalyzes oxidation of an aldehyde to an acid using NAD+ and water

A>Note: enzymes with this activity are involved in diverse metabolic pathways in various

C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

C:Keywords: alcohol metabolism, NAD, oxidoreductase

F:82-342/Domain, aldehyde dehydrogenase homology <ALDD>

F:288,322/Active site: Glu, Cys #status predicted

Query Match 67.8%; Score 1728.5; DB 1; Length 520.

Best Local Similarity 64.8%; Pred. No. 7.2e-106;

Matches 320; Conservative 81; Mismatches 92; Indels 1; Gaps 1;

QY 2 SHLPMTVPIKLPNGLEYEPTGLFINNKFPVSKQNTFEVINPSTFEFFICHYGRPDV 61

Db 24 SQAPLRVITLPGFTYEPTGLFINGEFVASKQKTPEDVINPSNEEKITTVYKAMEDDV 83

QY 62 EAVQAAADRAFSGNSWNGIDPIDRCKALRYLAFLIPQKQDVIASIETLDGKATSSPGD 121

Db 84 DEAVAAAKAFET-KWSIVEPEVRKALPNLADLVEKHQETLAATESMUNCKSLFCARSD 142

QY 122 VDLVINYLKSSAGFAUKIDGRMLDITGRTHFSTYTRQPIGVGGQIIPWNPFLIMMAWKIA 181

Db 143 VALSKYLRCSGWAKIKVGNVINDGKNHFTYSIKPELGVGGQIIPWNPFLIMMSWKITGP 202

QY 182 ALVTGNTVILKTAESTPLSALYSKYIPQAGIPPVIVINIVSGFGKIVVEATINHPKIKV 241

Db 203 ALATGNTVILKPAFTPLSALFASGLQAGALPA-VWNLIPSGVVGVEPSAHEVKKI 262

QY 242 APTGSTATGRHIVYSAAGALKKVIIGLQKSNIVFALAEUKAVVNIILGIYNSGRC 301

Db 263 APTGSTATGRHIMKVAADPVKKVILELQKSNIVFADADLDKAVKNIAFGIEYNSGRC 322

QY 302 CAGSRVTVESIVDKFEEFKASESIKVGDPDESTFGAOTSOMQNKILIKYVDIGKN 361

Db 323 CAGSRVTVESIVDKFEEFKASESIKVGDPDESTFGAOTSOMQNKILIKYVDIGKN 382

QY 362 EGATLITGGRIASGKGYFKPTVFGDVQKEIMRIVKEEIPGVVTVTKFKSADEVINMAND 421

Db 383 EGATLITGGRIASGKGYFKPTVFGDVQKEIMRIVKEEIPGVVTVTKFKSADEVINMAND 442
 QY 422 SEYGLAAGIHTSNINTALKVADRVNAGTVWINTYNDPHFHAVPFGGFGNAGSLGRENMSVDAL 481
 Db 443 SEYGLAAGIHTSNINTALKVADRVNAGTVWINTYNDPHFHAVPFGGFGNAGSLGRENMSVDAL 502
 QY 482 QNYLOVKAVRAKLD 495
 Db 503 SNTYOTKSVRIAD 516

RESULT 3

S60929

probable aldehyde dehydrogenase (NAD) (EC 1.2.1.4) YPL061w yeast (Saccharomyces cerevisiae)

N:Alternate names: protein IPE9w

C:Species: Saccharomyces cerevisiae

C>Date: 15-Feb-1996 #sequence_revision 31-Jan-1997 #text_change 04-Jun-2002

C:Accession: S60929

R:Wicker, F., Almer, A., Bussey, B., Fortin, N., Friesen, J.D., Hall, J., Storms, R.

submitted to the EMBL Data Library, October 1995

A:Reference number: S60921

A:Accession: S60929

A:Molecule type: DNA

A:Residues: 1-500 <WIN>

A:Cross references: EMBL:U08266; NID:G60722; PDB:AAK66041; PDB:1U796D; GSPDB:G

C:Genetics:

A:Gene: SGD:ALD6; MIPS:YPL061w

A:Cross references: SGD:S0005982; MIPS:YPL061w

A:Map position: 16L

C:Function:

A:Description: catalyzes oxidation of an aldehyde to an acid using NAD+ and water

A>Note: enzymes with this activity are involved in diverse metabolic pathways in vari

C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

C:Keywords: NAD, oxidoreductase

F:64-326/Domain, aldehyde dehydrogenase homology ALD6

F:272,306/Active site: Glu, Cys #status predicted

Query Match 56.9%; Score 1452; DB 1; Length 500;

Best Local Similarity 54.8%; Pred. No. 9.8e-88;

Matches 272; Conservative 92; Mismatches 128; Indels 4; Gaps 4;

QY 4 HLPMTVPIK LPNGLEYEPTGLFINNKFPVSKQNTFEVINPSTFEFFICHYGRPDV 60

Db 5 HPTAEPRVITLPGFTYEPTGLFINNKFPVSKQNTFEVINPSTFEFFICHYGRPDV 64

QY 61 VEFVAAAKAFET-KWSIVEPEVRKALPNLADLVEKHQETLAATESMUNCKSLFCARSD 120

Db 65 VEFVAAAKAFET-KWSIVEPEVRKALPNLADLVEKHQETLAATESMUNCKSLFCARSD 124

QY 121 DVDLVINYLKSSAGFAUKIDGRMLDITGRTHFSTYTRQPIGVGGQIIPWNPFLIMMAWKIA 180

Db 125 DVTIATNCLRDAAAYAKVNGRTINTGGMNFTTLEPVGVGQIIPWNPFLIMMAWKIA 184

QY 181 PALVTGNTVILKTAESTPLSALYSKYIPQAGIPPVIVINIVSGFGKIVVEATINHPKIKV 240

Db 185 PALAMNVGILKPAFTPLSALFASGLQAGALPA-VWNLIPSGVVGVEPSAHEVKKI 244

QY 241 VPTGSTATGRHIVYSAAGALKKVIIGLQKSNIVFALAEUKAVVNIILGIYNSGRC 299

Db 245 LAFTGSTATGRHIMKVAADPVKKVILELQKSNIVFADADLDKAVKNIAFGIEYNSGRC 304

QY 300 VYVAGSRVTVESIVDKFEEFKASESIKVGDPDESTFGAOTSOMQNKILIKYVDIGKN 358

Db 305 VYVAGSRVTVESIVDKFEEFKASESIKVGDPDESTFGAOTSOMQNKILIKYVDIGKN 364

QY 359 GKEPKATLITGGRIASGKGYFKPTVFGDVQKEIMRIVKEEIPGVVTVTKFKSADEVINM 418

Db 365 GKEPKATLITGGRIASGKGYFKPTVFGDVQKEIMRIVKEEIPGVVTVTKFKSADEVINM 424

QY 419 ARPSVYGLAAGIHTSNINTALKVADRVNAGTVWINTYNDPHFHAVPFGGFGNAGSLGRENMSVDAL 478

Db 425 ARPSVYGLAAGIHTSNINTALKVADRVNAGTVWINTYNDPHFHAVPFGGFGNAGSLGRENMSVDAL 484

QY 479 DALONYLOVKAVRAKL 494
Db 485 EYHAYTEVKAVRIKL 500

RESULT 4

A29055
aldehyde dehydrogenase (NAD) (EC 1.2.1.3) - *Emmericella nidulans*
C:Species: *Emmericella nidulans*, *Aspergillus nidulans*
C:Date: 30-Jun-1988 #sequence_revision 31-Jan-1997 #text_change 03-Jun-2002
C:Accession: A29055
R.Pickett, M.; Gwynne, D.I.; Buxton, F.P.; Elliott, R.; Davies, R.W.; Lockington, R.A.;
Gene 51, 217-226, 1987
A:Title: Cloning and characterization of the aldA gene of *Aspergillus nidulans*.
A:Reference number: A29055; MIM:87248080; PMID:3036652
A:Accession: A29055
A:Molecule type: DNA
A:Residues: 1-497 (PIR)
A:Cross-references: GR:M16197; NID:gl68010; PIDN:AAA33293.1; PID:gl68011
A:Experimental source: strain Glasgow F38c4
C:Genetics:
A:Gene: aldA
A:Map position: VIII
A:Introns: 21/2; 422/1
C:Function:
A:Description: catalyzes the oxidation of an aldehyde to an acid using NAD+ and water
A:Pathway: ethanol catabolism
A:Note: enzymes with this activity are involved in diverse metabolic pathways in various
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
C:Keywords: alcohol metabolism; NAD; oxidoreductase
F:57-318/Domain: aldehyde dehydrogenase homology <ALD>
F:264,298/Active site: 51u, Cys #status predicted

Query Match 55.4%; Score 1412.5; DB 1; Length 497;

Best Local Similarity 54.5%; Pred. No. 3.8e-85;
Matches 270; Conservative 88; Mismatches 132; Indels 5; Gaps 3;

QY 1 MSHLPMTVPKLPNGLEYEQPTGLFINNKEVFSKQNKTEFEVINSTEERIECHIEGKEDD 60
Db 1 MSDLPFTIETPV---IKYEQPLGLFINNEFVKVEGKTPAVINPSNEKVTISVHEATEKD 57
QY 61 VELAVGAADAPAFNSWN3IDPTDRGKALYPLAFILPQKIKVIVASIEFLNKAISSSGVDLVI 126
Db 58 VDVAVAAAAAF-EGSWRLPETPENRGKLLNNLANLFKNTDILAAVESLDNGKATSMARVTSACAS 121
QY 121 DVDLVINYLKSSAGFADKIDGRMTDGTGRTHFSYTKRQPLGVCGGIIIPWNPFLMVAWKIA 180
Db 117 DLANSIGCLRYAGWADKTHGQTTDTNPETLTITRPEPVGVCCGIIIPWNPFLMWSWKIG 176
QY 181 PALVTGNTVVLKTAESTPLSALYVSKYIPQAGIPPGVINIVSGFGKIVVEATINHPKIRK 240
Db 177 PAVAAAGNTVVLKTAQTPLSALYAAKLIKEAPFAGVINIVSGFGRTAGAAISSHMDIDK 236
QY 241 VAFTGSTATGTHIYQSA-AGLKVVITLGGKSNIVFADAELKKAQVNLILGIYNSGE 299
Db 237 VAFTGSTLVLGTTILQAAKSNLKVVITLGGKSNIVFDDADILNALSMAWKFIFPNHGQ 296
QY 300 VCCAGSRVYVESIYDKFIEFFKAASESIKVSGIPFDETFQGAQTSQMGLNKLKLYVDIG 359
Db 297 CCCAGSRILVQEGIYDKIVAKFKKAKNKNVGPFEQDTFQGVQVQLQDFRIMEYING 356
QY 360 KNEGATLTGGERIGSGGYFIKTPVFGDVKEDMRIVKEEIPGVVTVTKFSADEVINMA 419
Db 357 KKAGATVATGDKRGNBGFILQPVFTLDVTSUMKIAQEEIFGFPVTVLQKPKDVAEAIKIG 416
QY 420 NDSEYGLAAGTHTSNITALKVADRVNAGTVWINTYNDFFHVAVPFGGNASGLGKREMSVD 479
Db 417 NSTDYGAAAVHTKNVNTAIRVSNALKAGTVWNNYNNIYSTQAPFGGFGKSGLGRGLG 476
QY 480 ALONYLOVKAVRAKL 494
Db 477 ALENTYQIKTVHYRL 491

RESULT 5

S43114
aldehyde dehydrogenase (NAD) (EC 1.2.1.3) - fungus (*Cladosporium herbarum*)
C:Species: *Cladosporium herbarum*
C:Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 03-Jun-2002
C:Accession: S43114
F.Achtatz, G.; Oberkofler, H.; Simon, B.; Lechenauer, F.; Unger, A.; Kandler, D.; Pril
submitted to the EMBL Data Library, March 1994
A:Description: Molecular characterization of allergens of *Cladosporium herbarum* and A
A:Reference number: S43108
A:Accession: S43114
A:Molecule type: mRNA
A:Residues: 1-496 (ACH>
A:Cross-references: EMBL:X74228; NID:3457624; PIRN:AA57072.1; PIR:3457625
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
C:Keywords: NAD; oxidoreductase
F:57-317/Domain: aldehyde dehydrogenase homology <ALD>

Query Match 54.9%; Score 1399; DB 2; Length 496;

Best Local Similarity 53.8%; Pred. No. 2.9e-84;
Matches 263; Conservative 89; Mismatches 135; Indels 2; Gaps 3;

QY 7 TVPIKLPNGLEYEQPTGLFINNKEVFSKQNKTEFEVINSTEERIECHIEGKEDDVEAVQ 66
Db 3 SVQLETPHSGKYEQPTGLFINNEFVKVEGKTEFDVNPDESIVTVHEATEKDVIDIAVA 62
QY 67 AADPAFNSWN3IDPTDRGKALYPLAFILPQKIKVIVASIEFLNKAISSSGVDLVI 126
Db 63 AARQAF-EGSWRLPETPENRGKLLNNLANLFKNTDILAAVESLDNGKATSMARVTSACAS 121
QY 1-7 NYLKSSAGFADKIDGRMTDGTGRTHFSYTKRQPLGVCGGIIIPWNPFLMVAWKIAPALVTG 186
Db 122 GCLRYAGWADKTHGQTTDTNPETLTITRPEPVGVCCGIIIPWNPFLMVAWKIAPALVTG 181
QY 187 NTVVLKTAESTPLSALYVSKYIPQAGIPPGVINIVSGFGKIVVEATINHPKIRKVAFTGS 246
Db 184 NTVVLKTAESTPLSALYVSKYIPQAGIPPGVINIVSGFGKIVVEATINHPKIRKVAFTGS 241
QY 247 TATGRTHIYQSA-AGLKVVITLGGKSNIVFADAELKKAQVNLILGIYNSGEVCCAS 305
Db 242 TVGRTILKAAASNLKVVITLGGKSNIVFDDADILNALSMAWKFIFPNHGQVCCAS 301
QY 306 PVYVESIYDKFIEFFKAASESIKVSGIPFDETFQGAQTSQMGLNKLKLYVDIGKNEFAT 365
Db 302 RVYVESIYDKFIEFFKAASESIKVSGIPFDETFQGAQTSQMGLNKLKLYVDIGKNEFAT 361
QY 366 LITGGERIGSGGYFIKTPVFGDVKEDMRIVKEEIPGVVTVTKFSADEVINMANDEYV 425
Db 362 VETGSRKRGKGYFIETPFSNVTEDMKIVKEEIPGVVTVTKFSADEVINMANDEYV 421
QY 426 LAAGTHTSNITALKVADRVNAGTVWINTYNDFFHVAVPFGGNASGLGKREMSVDALONYL 485
Db 422 LAAVHTKNLNTALEVSNALKAGTVWNNYNNIYSTQAPFGGFGKSGLGRGLG 481
QY 486 QVKAVRAKL 494
Db 482 QTKTVSIRL 490

RESULT 6

S43108
aldehyde dehydrogenase (NAD) (EC 1.2.1.3) - *Alternaria alternata*
C:Species: *Alternaria alternata*
C:Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 03-Jun-2002
C:Accession: S43108
F.Achtatz, G.; Oberkofler, H.; Simon, B.; Lechenauer, F.; Unger, A.; Kandler, D.; Pril
submitted to the EMBL Data Library, March 1994
A:Description: Molecular characterization of allergens of *Cladosporium herbarum* and A
A:Reference number: S43108
A:Accession: S43108
A:Molecule type: mRNA
A:Residues: 1-495 (ACH>

C:Function:

A:Description: catalyzes the oxidation of an aldehyde to an acid using NAD⁺ and water
 A:Pathway: ethanol catabolism
 C:Superfamily: aldehyde dehydrogenase (NAD⁺); aldehyde dehydrogenase homology
 C:Keywords: alcohol metabolism; homotetramer; liver; mitochondrion; NAD⁺; oxidoreductase
 F:1-19/Domain: transit peptide (mitochondrion) #status predicted <TM>
 F:23 51%/Product: aldehyde dehydrogenase (NAD⁺) 2 #status predicted <MAL>
 F:77-341/Domain: aldehyde dehydrogenase homology <ALDH>
 F:287,321/Active site: Glu, Cys #status predicted
 F:474/Binding site: NAD (Cys) #status predicted

Query Match 50.0%; Score 1275.5; DB 1; Length 519;
 Best Local Similarity 51.0%; Pred No. 4e-76;
 Matches 253; Conservative 84; Mismatches 146; Indels 13; Gaps 6,

QY 7 TVPIKLPNGLEYEQP-----TGLFINNKVPKSKQNKTEVINPSTEEIECHLYCKEEDVE 62
 Db 24 TSAVAPN-----HQPEVFCNQIFINNEWHDAVSRKKTFTVNPSTGEVICQVAGCKEDVD 79
 QY 63 EAVQAADRAFNSGS--WNIGIDPDKGKALYKLAFLLEKQKVIASTETLDNGKA-ISSSKG 120
 Db 80 KAVKARAARAFQSGSPRMKASDRGRILYPLADLIERPTVIALETLDNGKPYVISLY 139
 QY 121 DVDLVNLYLKSSAGFADKIDGRMIDTPTSTFSTYKPKPLGVCQGIIPWNPFLLMWAKKIA 180
 Db 140 DLDMLVKLPYVYAGWADKYHCKTIPIDGDFEFTYHRRPVGVCQGIIPWNPFLLMWAKKLG 199
 QY 181 PALVTGNTVVLKTAESTPLSALYVSKYIPQAGIPPGVINIVSGPKIIVVATINHPKIKK 240
 Db 200 PALATGNVVMVKVAEQTPTLTALYVANLKEAGFPFGVNVIVPGRGPTAGAAIASHEGVDK 259
 QY 241 VAFVSTATGRHIYGSAA--AALKKVFLDGSKSFNIFALDAELKKAVQNIILGIYVNSG 298
 Db 260 VAFVSTEVG--HLIVAAAGSSNLKVKVLELDGSKSFNIIIMSLAIDMWAVEVAHAFPNQG 318
 QY 299 EYVQAGSRVYVEESLYDKFIEEFKRAASEIKVGDFEESTFGQAGISQMIINKILKYVDI 358
 Db 319 QQQAGSRLEFGVENVYDEFEVERSVAFKSRVGVNPGNFSPTFGQPVQDETQPKKILGYIKS 478
 QY 359 GNGEATLITGCEPLGSKGYFKTPVFGVWKEHMPVKEEIEFGPVVTVKPKSADRVINM 418
 Db 379 GQGERAKILGGGAAAGRQYFQPIVFGVWKEHMPVMTIAKEIFGPGVMQILAFKTEEVVGP 448
 QY 419 ANDSEYGLAAGIHISNITALKVAQVNAQTWINTYNDHFHVPFGSGFNASSIGSGEMSV 478
 Db 439 ANUSKYLAAAVFTKDLKANYLSALCAGIWINVYIVFCAUSPFGSKYKMSGSRPELGE 498
 QY 479 DALQNYLQVKAVRAKL 494
 Db 499 YGLQAYTEVKTVTKV 514

RESULT 9

S03564

aldehyde dehydrogenase (NAD) (EC 1.2.1.3) 2 precursor, mitochondrial - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 28-Feb-1990 #sequence_revision 31 Jan-1997 #text_change 03-Jun-2002

C:Accession: S03564, A27713; S12903; S17432

R:Farres, J.; Guan, K.L.; Weiner, H.

Eur J Biochem 180, 67-74, 1989

A:Title: Primary structures of rat and bovine liver mitochondrial aldehyde dehydrogenase

A:Reference number: S03564; MIM:14210865; PMID:2540603

A:Accession: S03564

A:Molecule type: mRNA

A:Residues: 1-519 <FAP>

A:Cross-references: EMBL X14977; NID:955604; PDB:CAA33101 1; PDB:955605

A:Experimental source: strain Sprague Dawley; liver

R:Farres, J.; Guan, K.L.; Weiner, H.

Biochem Biophys Res Commun 150, 1083-1087, 1988

A:Title: Sequence of the signal peptide for rat liver mitochondrial aldehyde dehydrogenase

A:Reference number: A27713; MIM:88134217; PMID:3342060

A:Accession: A27713

A:Molecule type: mRNA

A:Residues: 1-519 <FAL>

A:Cross-references: GB M36930; NID:9302847; PDB:AAA071913; PDB:9556048

R:Diwan, J.; Pallwal, K.; Kaftan, E.; Bawa, R.

FEBS Lett 273, 215-218, 1990

A:Title: A mitochondrial protein fraction catalyzing transport of the K(+) analog Tl(+)

A:Reference number: S13903; MIM:4162184; PMID:164808

A:Accession: S12903

A:Molecule type: protein

A:Residues: 327-430 <EDW>

R:Jeng, J.; Weiner, H.

Arch Biochem Biophys 389, 214-222, 1991

A:Title: Purification and characterization of catalytically active precursor of rat l

A:Reference number: S17432; MIM:4174540; PMID:164808

A:Accession: S17492

A:Molecule type: protein

A:Residues: 1-19 <FEN>

C:Genetics:

A:Genome: nuclear

C:Complex: homotetramer

C:Function:

A:Description: catalyzes the oxidation of an aldehyde to an acid using NAD⁺ and water

A:Pathway: ethanol catabolism

A:Note: enzymes with this activity are involved in diverse metabolic pathways in vari

C:Superfamily: aldehyde dehydrogenase (NAD⁺); aldehyde dehydrogenase homologyC:Keywords: alcohol metabolism; homotetramer; liver; mitochondrion; NAD⁺; oxidoreducta

F:1-19/Domain: transit peptide (mitochondrion) #status predicted <TM>

F:20-519/Product: aldehyde dehydrogenase (NAD⁺) 2 #status predicted <MAL>

F:77-341/Domain: aldehyde dehydrogenase homology <ALDH>

F:287,321/Active site: Glu, Cys #status predicted

F:474/Binding site: NAD (Cys) #status predicted

Query Match 50.0%; Score 1274.5; DB 1; Length 519;

Best Local Similarity 50.8%; Pred No. 4.7e-76;

Matches 252; Conservative 85; Mismatches 146; Indels 13; Gaps 6;

QY 7 TVPIKLPNGLEYEQP-----TGLFINNKVPKSKQNKTEVINPSTEEIECHLYCKEEDVE 62

Db 24 TSAVAPN-----HQPEVFCNQIFINNEWHDAVSRKKTFTVNPSTGEVICQVAGCKEDVD 79

QY 63 EAVQAADRAFNSGS--WNIGIDPDKGKALYKLAFLLEKQKVIASTETLDNGKA-ISSSKG 120

Db 80 KAVKARAARAFQSGSPRMKASDRGRILYPLADLIERPTVIALETLDNGKPYVISLY 139

QY 121 DVDLVNLYLKSSAGFADKIDGRMIDTPTSTFSTYKPKPLGVCQGIIPWNPFLLMWAKKIA 180

Db 140 DLDMLVKLPYVYAGWADKYHCKTIPIDGDFEFTYHRRPVGVCQGIIPWNPFLLMWAKKLG 199

QY 181 PALVTGNTVVLKTAESTPLSALYVSKYIPQAGIPPGVINIVSGPKIIVVATINHPKIKK 240

Db 200 PALATGNVVMVKVAEQTPTLTALYVANLKEAGFPFGVNVIVPGRGPTAGAAIASHEGVDK 259

QY 241 VAFVSTATGRHIYGSAA--AALKKVFLDGSKSFNIFALDAELKKAVQNIILGIYVNSG 298

Db 260 VAFVSTEVG--HLIVAAAGSSNLKVKVLELDGSKSFNIIIMSLAIDMWAVEVAHAFPNQG 318

QY 299 EYVQAGSRVYVEESLYDKFIEEFKRAASEIKVGDFEESTFGQAGISQMIINKILKYVDI 358

Db 319 QQQAGSRLEFGVENVYDEFEVERSVAFKSRVGVNPGNFSPTFGQPVQDETQPKKILGYIKS 478

QY 359 GNGEATLITGCEPLGSKGYFKTPVFGVWKEHMPVKEEIEFGPVVTVKPKSADRVINM 418

Db 379 GQGERAKILGGGAAAGRQYFQPIVFGVWKEHMPVMTIAKEIFGPGVMQILAFKTEEVVGP 448

QY 419 ANDSEYGLAAGIHISNITALKVAQVNAQTWINTYNDHFHVPFGSGFNASSIGSGEMSV 478

Db 439 ANUSKYLAAAVFTKDLKANYLSALCAGIWINVYIVFCAUSPFGSKYKMSGSRPELGE 498

QY 479 DALQNYLQVKAVRAKL 494

Db 499 YGLQAYTEVKTVTKV 514

RESULT 10

DEH02
 aldehyde dehydrogenase (NAD) (EC 1.2.1.3) 2 precursor, mitochondrial [validated] - human
 N:Alternate names: aldehyde dehydrogenase E2; aldehyde dehydrogenase I
 C:Species: Homo sapiens (man)
 C:Date: 31-Mar-1988 #sequence revision 17-Nov-1995 #text change 04-Jun-2002
 C:Accession: A29975; S00804; A23503; A27509; A26743; I39433; I39430
 R:Hsu, L.C.; Benda, P.E.; Yoshida, A.
 Genomics 2, 57-65, 1988
 A:Title: Genomic structure of the human mitochondrial aldehyde dehydrogenase gene.
 A:Reference number: A29975; MUID:88256152; PMID:2838413
 A:Accession: A29975
 A:Molecule type: DNA
 A:Residues: 1-517 <HSU1>
 A:Cross-references: GB:M26760; NID:q178397; PIDD:AAA51694.1; PIDD:q178398
 R:Braun, T.; Bober, E.; Singh, S.; Agarwal, D.P.; Goedde, H.W.
 FEBS Lett. 233, 340, 1988
 A:Reference number: S00804
 A:Accession: S00804
 A:Molecule type: mRNA
 A:Residues: 1-6, 'APA', 10, 'P', 13-49, 'UPA'
 A:Cross-references: EMBL:X05409; NID:928605
 A:Note: Contribution to A26743
 R:Hempel, J.; Kaiser, R.; Joernvall, H.
 Eur. J. Biochem. 153, 13-28, 1985
 A:Title: Mitochondrial aldehyde dehydrogenase from human liver. Primary structure, difference
 A:Reference number: A23503; MUID:86055846; PMID:4065146
 A:Accession: A23503
 A:Molecule type: protein
 A:Residues: 1-517 <HEM>
 A:Note: The sequence shown is presumably that of the mature protein, however, the number
 R:Braun, T.; Bober, E.; Singh, S.; Agarwal, D.P.; Goedde, H.W.
 Nucleic Acids Res. 15, 3179, 1987
 A:Title: Isolation and sequence analysis of a full length cDNA clone coding for human mi
 A:Reference number: A27509; MUID:87174836; PMID:3562250
 A:Accession: A27509
 A:Molecule type: mRNA
 A:Residues: 1-6, 'APAWA', 10, 'P', 12, 'VS', 15, 'RHPRG', 21, 27-79, 'REGPRG', 86-336, 'V', 338-517
 A:Cross-references: EMBL:X05409; NID:928605
 A:Note: This sequence is revised in reference S00804
 R:Hsu, L.C.; Tani, K.; Fujiyoshi, T.; Kurochi, K.; Yoshida, A.
 Proc. Natl. Acad. Sci. U.S.A. 82, 3771-3775, 1985
 A:Title: Cloning of cDNAs for human aldehyde dehydrogenases 1 and 2.
 A:Reference number: I39432; MUID:85216574; PMID:2987944
 A:Accession: I39432
 A:Molecule type: mRNA
 A:Status: preliminary
 A:Residues: 119-336, 'V', 338-517 <HSU2>
 A:Cross-references: GR:K03001; NID:q178395; PIDD:AAH59500.1; PIDD:q178396
 A:Note: thirty-three tryptic peptides were also sequenced
 R:Yoshida, A.; Ikawa, M.; Hsu, L.C.; Tani, K.
 Alcohol 2, 103-106, 1985
 A:Title: Molecular abnormality and cDNA cloning of human aldehyde dehydrogenases.
 A:Reference number: I39431; MUID:85252089; PMID:4015823
 A:Accession: I39431
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 119-336, 'V', 338-517 <YOS>
 A:Cross-references: GB:M26760; NID:q178397; PIDD:AAA51694.1; PIDD:q178398
 R:Agarwal, D.P.; Goedde, H.W.
 Isozymes Curr. Top. Biol. Med. Res. 16, 21-48, 1987
 A:Title: Human aldehyde dehydrogenase isozymes and alcohol sensitivity.
 A:Reference number: I39430; MUID:87774033; PMID:4610592
 A:Accession: I39430
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA
 A:Residues: 214-215, 'S', 217, 'K', 219-246, 'P', 248-436, 'V', 438-426, 'EVGIMRGGWHSQQFHHVKG'
 A:Cross-references: GB:M54931; NID:q178391; PIDD:AAA52625.1; PIDD:q178392
 A:Note: the sequence is misidentified as aldehyde dehydrogenase 1
 C:Genetics:
 A:Gene: GDB:A1DH2
 A:Cross-references: GDB:112608, OMIM:100650
 A:Map position: 12q24.2-12q24.2
 A:Insertions: 4874, 7474, 12674, 14724, 1843, 22274, 26674, 30074, 34124, 37974, 41474, 45074, 48774, 52474
 C:Complex: homotetramer
 C:Function:
 A:Description: catalyzes the oxidation of an aldehyde to an acid using NAD+ and water
 A:Pathway: ethanol catabolism
 A:Note: enzymes with this activity are involved in diverse metabolic pathways in vari
 C:Superfamily: aldehyde dehydrogenase (NADP+), aldehyde dehydrogenase homology
 C:Keywords: alcohol metabolism; homotetramer; liver; mitochondrion; NAD; oxidoreducta
 F:1-17/Domain: transit peptide (mitochondrion) #status predicted -SUS
 F:18-517/Product: aldehyde dehydrogenase (NADP+) #status experimental -MAF
 F:75-339/Domain: aldehyde dehydrogenase homology -ALDO
 F:211-291/Domain: NAD binding #status predicted -NAD
 F:285/Active site: site #status predicted
 F:319/Active site: Cys #status experimental
 F:472/Binding site: NAD (Cys) #status predicted

Query Match 49.2% Score 1254.5 DB 1 Length 517
 Best Local Similarity 50.4% Pred. No. 1 to 74
 Matches 249; Conservative 86; Mismatches 149; Indels 11; Gaps 5
 QY 7 TVPIKLPNGLEYEQ-----TGLPIINNKVPISKONKPEVININSTEETICHLYNREHIVE 62
 DB 22 TQAVFAFN - QQTEVECNQFINNEWHDAVERKTTTVRISGEVTVQAEIRKEIVD 77
 QY 63 EAVQAAADRAFSNCS-WNGIDDPIDRCALYLAELIDQDKVIASITETLNKKA -LSSSRG 120
 DB 78 KAVKAAKAAAFGLSSPKWKKMLASRSHLLNKAHLERFETLAALEFLNKKPVVSYLV 147
 QY 121 DVDLVINYLKSSAGFAURKIDGRMIDGRTHFSYTRKQPLAVGQITPNNFELIMKAWKIA 180
 DB 138 DLMVMVLEKLYYAGWAKYKUGKILFGLGLFSAIKRDLWVWQQTIVWNLLEMGAWKAG 197
 QY 181 PALVGNIVVKEIKESPLSALYVSKYDPAWGLPEVINYSEKIVVVALEINRERK 240
 DB 198 PALATNNVVMKVAIVTPIALYVANGKEAVPTPEVNVN;VNGTPTPA-MAALAHVHP 257
 QY 241 VAFSTATGSETHYGA-AAAGFPVTHETGPTTHIVIAALAPFAVNIETIYNSPE 299
 DB 258 VAFSTGTEICRVLTQVAAAGSNLKKVTLGLGKSNLINSIDALMWAVQAHLFENQSG 317
 QY 359 VQFASFPVVEESTYHFFEFKKAANSELYKVDPPESLFGAQUSGMENFHLKYVDIG 359
 DB 318 QGASGSKTFVQEDTDLAVLRGVARAKSKVGNRFRERKELQGVQVRLGQIKKIDYIN 372
 QY 360 KNHGATLTGGERLQSKGYFKPTVPRVAKEMIRVKEEFEDGVVETKPSADEVINMA 419
 DB 378 KQEGAKLGGGGLAIBRGYFLOPTVEGVQVQKMTIAKEITFGVWMLIKPKLEEVVIRA 447
 QY 420 NUSEYGLAAGIHISININIAKVAADVKNACTVMTINYNIIHVAIVPEFGNASKREMSVD 479
 DB 448 NNSTYGLAAAVFTKDLKANYLSQALQACTVWNVYVWKAUSDFPQYKMSGSGRELEY 497
 QY 480 ALQNYLOVKAVRAKL 494
 DB 498 GLQAYTEVKTVTKV 512

RESULT 11

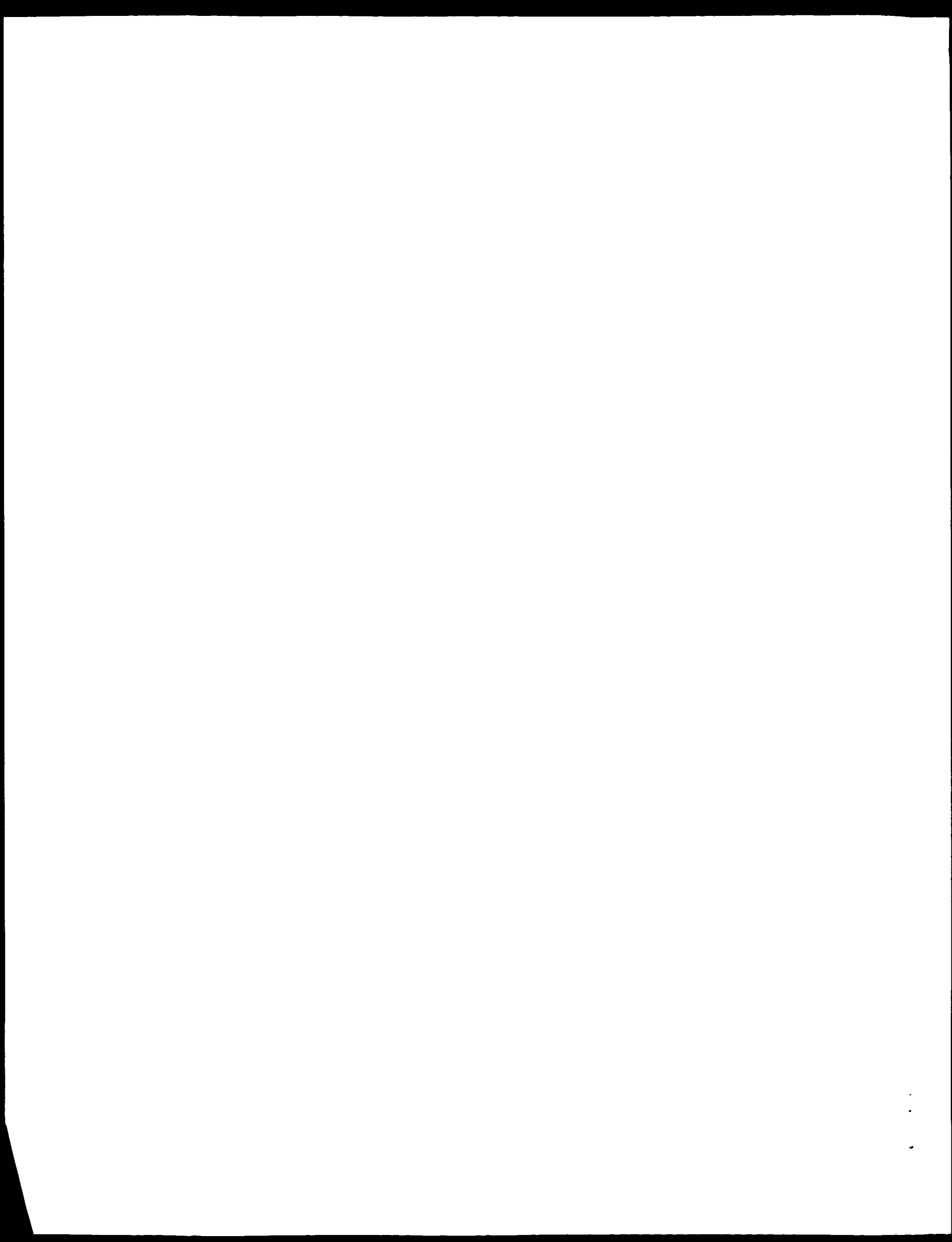
B69614
 aldehyde dehydrogenase dbas - Bacillus subtilis
 C:Species: Bacillus subtilis
 C:Date: 05-Dec-1997 #sequence revision 05-Dec-1997 #text change 20-Jun-2000
 C:Accession: B69614
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azavedo, V.; Bor
 C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caddow, B.; Capuano, V.; Carter, N.M.;

N: Alternate names: aldehyde dehydrogenase homolog
 C: Species: Zea mays (maize)
 C: Entry: J03983
 C: Accession: T03983
 R: Cui, X.; Wise, R.P.; Schnable, P.S.
 Science 272, 1334-1336, 1996
 A: Title: The r12 nuclear restorer gene of male-sterile T-cytoplasm maize.
 A: Reference number: 215177; MUID: 96243131; PMID: 8650543
 A: Accession: T03983
 A: Status: preliminary; translated from GB/EMBL/DBJ
 A: Molecule type: mRNA
 A: Residues: 1-549 <CUI>
 A: Cross-references: EMBL: J03983, NID: 96243131, FID: 96243131
 A: Experimental source: strain B73
 C: Genetics:
 A: Map position: 9
 A: Note: r12
 C: Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 F: 107-370/domain: aldehyde dehydrogenase homology <ALD>

Query Match 48 38; Score 1232.5; DB 2; Length 549;
 Best Local Similarity 48.1%; Pred. No. 2.9e-73;
 Matches 244; Conservative 51; Mismatches 143; Indels 23; Gaps 5;

Qy	7	TVPIK-----LPNGLE-----YEOP-----TGLFINNKFVPSKQNKTFEVIINPS	45
Db	33	TVPADGMHPLPCVLOPFSTAAAVFPPITPSVHVNYTKLLINGNEVDSASGKTFPTLPP	92
Qy	46	TEETICHIEKREDVVEAVQAAKAFNSGNGIDPDKKALYKLAELIEQDKDVAS	105
Db	93	1GEVIAHVAEGDAEDINRAVAAKAKAFDESPNEKMTAYEKSKILLKFAADLIEKHNDLAA	152
Qy	106	IELDNGKAT-SSSRGVDLAVINYLKSSAGFAKIDGPMIDGPTHFSTYKRPQIGVCGQ	164
Db	153	LETWNGKPYEQAAQTEVPVAFIMBYAATWAKTHSIVPAFGPHHVGLIHEPISVAGQ	212
Qy	165	IIPWNPFLLMWANKIAPALVTGNTVLKTAESTPLSALYVSKYIPOAGIPPGVINIVSGF	224
Db	213	IIPWNPFLMYAWKVPALACGNTLVLTAEQTPLSALYISKLLHENGLEPGVWVWSGF	272
Qy	225	GKIVVEATNHPIKKVAFITGSTATGPHIYQSA-AGLKKVTLLELGGKSPNIVFADAEK	283
Db	273	GPTAGAAASHMDVKIAFTGSTDTGKIILELAASNKLTVTLELGGKSPFIIMDDADV	332
Qy	284	KAVENITIGIVYNSGEVCCAGSRVYVEESTYKFIIEFKAASESIKVGDPPDESTFQGAQ	343
Db	333	HAVELAHFALFFNGUCCAGSPTFVHEKVFDEFKAKAPALKPVVCPPEKSGVEGHPQ	392
Qy	344	TSQMLNKLKYYVDIGKNEGATLITGGERIGSGGYFIKPTVFGDVKEDMRIYKEEIPGPV	403
Db	393	IDDEQFNKILLYIPYGVGGGATLVTSRQFQJHKGFIQPTIFSDVLGSKMKIAEELFGPV	452
Qy	404	VTVTKEKSADEVINWANDSEYGLAAGIHTSNITALKVADRVNAGTVWINTYNDEHHAVP	463
Db	453	QSILKFKDLNEVTKRNASQYGLAAGVTINSLDTANTLTFAURAGIVWVNCDFVDAATP	512
Qy	464	EGGFNASLGRFMSVDALQNYLQVKAV	490
Db	513	EGGYKMSGIGREKGVDSILKNYLQVKAV	539

Search completed: June 24, 2003, 10:18:29
 Job time : 20.9426 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Pur 20: June 24, 2003, 10:02:55 ; Search time 0.69979 seconds
(without alignments)
2059,300 Million cell updates/sec

Title: US-09-830-751-2
Perfect score: 2550
Sequence: 1 MSHLPMTVPTKLPNGLEVEQ

MSVDALQNYLQNKAVPAKLD 495

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 5%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2538	99.5	519	1 DH44 YEAST	P46367 saccharomyc
2	1695	66.5	519	1 DH45 YEAST	P40047 saccharomyc
3	1455	57.1	497	1 DH46 ASPERG	P41751 aspergillus
4	1452	56.9	499	1 DH46 YEAST	P54115 saccharomyc
5	1412.5	55.4	497	1 DH46 EMENI	P08157 emericella
6	1399	54.9	496	1 DH46 CLARE	P40108 cladosporiu
7	1362.5	53.4	495	1 DH46 ALITAL	P42041 alternaria
8	1275.5	50.0	519	1 DH46 MOUSE	P47734 mus musculi
9	1274.5	50.0	519	1 DH46 RAT	P11894 rattus norv
10	1261.5	49.5	500	1 DH46 MESAI	P11778 mesocricetu
11	1253.5	49.2	517	1 DH46 HUMAN	P05041 homo sapien
12	1248.5	49.0	517	1 DH45 HUMAN	P30837 homo sapien
13	1244.5	48.8	499	1 DH45 CHICK	O93344 gallus gall
14	1240.5	48.6	520	1 DH46 BOVIN	P20000 bos taurus
15	1230	48.2	499	1 DH46 HUMAN	O94788 homo sapien
16	1223.5	48.0	499	1 DH46 RAT	O63639 rattus norv
17	1222.5	47.9	499	1 DH46 MOUSE	O62148 mus musculu
18	1220.5	47.9	500	1 DH46 HORSE	P12762 equus cabal
19	1219	47.8	500	1 DH46 AGART	O74187 atalapha L.
20	1218.5	47.8	512	1 DH46 HUMAN	P47895 homo sapien
21	1212.5	47.5	500	1 DH46 CHICK	O57453 gallus gall
22	1192.5	46.8	497	1 DH46 FRUIT	O57643 ancyliu
23	1182.5	46.4	500	1 DH46 HUMAN	P00352 homo sapien
24	1181.5	46.3	500	1 DH46 SHEEP	P51977 ovis aries
25	1179.5	46.3	500	1 DH46 PAT	P51647 rattus norv
26	1176.5	46.1	500	1 DH46 HORSE	P15437 equus cabal
27	1175.5	46.1	500	1 DH46 MOUSE	P24549 mus musculu
28	1174.5	46.1	500	1 DH46 BOVIN	P48644 bos taurus
29	1169.5	45.9	500	1 DH46 PAT	P13601 rattus norv
30	1155.5	45.3	501	1 DH46 MACPP	O29450 macroscelid
31	1149.5	45.1	501	1 DH46 FURID	O28499 elephanu
32	1141	44.7	498	1 DH46 LEITA	O25417 leishmania
33	1101.5	43.2	495	1 CPOM OCTDO	P30841 octopus dof

34 1094 44.1 902 1 FTDH PAT
35 1095 42.9 494 1 CPOM OMMSL
36 1092 42.8 506 1 DH42 YEAST
37 1067 41.8 902 1 FTDH HUMAN
38 1057 41.5 506 1 DH43 YEAST
39 1034.5 40.6 490 1 DHAB BAGSU
40 1006 39.5 511 1 DHAY YEAST
41 948.5 37.2 499 1 FEAB EYLI
42 942.5 37.0 487 1 DHAB RLIME
43 936 36.7 497 1 DHAB STIOL
44 930 36.5 512 1 ALDB ECOLI
45 919.5 36.1 506 1 DHAB VTRCH

ALIGNMENTS

RESULT 1
DH44 YEAST
ID DH44 YEAST STANDARD: PRT: 519 AA.
AC P46367: Q08898;
DT 01-NOV-1995 (rel. 32, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Potassium-activated aldehyde dehydrogenase, mitochondrial precursor
DE (EC 1.2.1.3) (K(+)-activated acetaldehyde dehydrogenase) (K(+)-ACDH).
GN ALP4 OP ALP7 OP ALP2 OP YOP374W OP 06730.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetia; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Delius H., Hobling U., Hofmann B.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 25-65.
RX MEDLINE=91113163; PubMed=1989592;
RA Chalmers P.M., Keen J.N., Fowson C.A.;
RT "Comparison of benzyl alcohol dehydrogenases and benzaldehyde
RT dehydrogenases from the benzyl alcohol and mandelate pathways in
RT Acetobacter carotenicus and from the TOL-plasmid-encoded toluene
RT pathway in Pseudomonas putida. N-terminal amino acid sequences, amino
RT acid compositions and immunological cross-reactions.";
RL Biochem. J. 273:99-107(1991).
RN [3]
RP SEQUENCE OF 25-34 AND 378-386.
RX MEDLINE=9295278; PubMed=9150920;
RA Larsson T., Norbeck P., Karlsson H., Karlsson K.-A., Blomberg A.;
RT Identification of two dimensional gel electrophoresis resolved yeast
RT proteins by matrix-assisted laser desorption/ionization mass
RT spectrometry.";
RL Electrophoresis 18:418-423(1997).
RN [4]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RX MEDLINE=9340498; PubMed=925847;
RA Foster W.H., Meaden P.G., Dickinson F.M., Midgley M.;
RT Identification and disruption of the gene encoding the F(1)F(0)-driven
RT acetaldehyde dehydrogenase of Saccharomyces cerevisiae.";
RL FEMS Microbiol. Lett. 164:39-44(1998)
RN [5]
RP CATALYTIC ACTIVITY: An aldehyde + NAD(+) -> an acid + NADH.
RN [6]
RP SUBCELLULAR LOCATION: Mitochondrial matrix.
RN [7]
RP SIMILARITY: BELONGS TO THE ALDH4 DEHYDROGENASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb.sib.ch/annouco/>)
CC or send an email to license@isb-sib.ch
CC -----
CC EMBL; 275282; CAA99705.1; -

[illegible]


```

QY 180 APALVTGNTVTLKTAESTPLSALYSKYIPQAGIPPGVIVINVSSEKIVVEAINTNPKIK 239
DB 200 GPALATCINTVILKPAETTHLSALPASQAQAGIPAGVNVLLTSGSVGVGSRISAIIDVK 259
QY 240 KVAFTGSTATGPHIYQSAAGIKKVTLEQSKSPNIVFAIAIKKAVQNIILGLIYNSSE 299
DB 260 KIAPTGSTATGPHIMKVAAGVKKVTLLEAGKSPNIVFAIADILKAVKNIAPGTEYNSSE 319
QY 300 VCCAGSPVYVESIYDKLEBEFKAASBSIKVGDPEFDESHIPQAGVISMJLNKILKYVDIG 359
DB 320 VCCAGSRIVQDVTVEEVLQKDKYTBSLKVGDPDEEVFGAGTSDKQLRKILDYVDA 379
QY 360 KNGATLITGGERLASKGYFIKPTVFGVSKEDMBIVKEEIPGVPVITVKFSADENVMA 419
DB 380 KSGAPVITGGAPHSKGYFVKPTVFAVDKEDMBIVKEEIPGVPVITVKFSADENVMA 439
QY 420 NDSYGLAAGIHTSINTALKVADRNVAGTVMINTYNDPHHAFVFGFNGSAGIGREMSVD 479
DB 440 NDSYGLAAGIHTNDINKAVIVSKVRKAGIWMINTYNNPHQNVKGGFGGSGIPGEMGEA 499
QY 480 ALQNYLVKAVKAKLD 495
DB 500 ALSNTQTQKSVRIAD 515

RESULT 3
DHAL_ASPNG STANDARD: PRT: 497 AA.
AC P41751;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ALdehyde dehydrogenase (EC 1.2.1.3) (ALDH).
GN ALDA.
OS Aspergillus niger.
OC Eukaryota, Fungi, Ascomycota, Pezizomycotina, Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RP SEQUENCE FROM N A
PX MEDLINE=90108706; PubMed=2606357;
RA O'Connell M J, Kelly J M.;
FT "Physical characterization of the aldehyde dehydrogenase-encoding
RL gene of Aspergillus niger";
RL Gene 84:173-180(1989)
CC -1- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O -> an acid + NADH.
CC -1- PATHWAY: Ethanol utilization; second step.
CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC
CC This SWISS-PROT entry is copyright © 1993. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announc/
CC or send an email to license@sib-sb.ch).
CC
CC EMBL; M32351; AAA87596.1;
CC HSSP; P05091; LCM3.
CC InterPro: IPR02086; Aldehyde_dehydr
CC Pfam: PF00171; aldehyd; 1.
CC PROSITE: PS00070; ALDEHYDE_DEHYDR_GYS; 1
CC PROSITE: PS00687; ALDEHYDE_DEHYDR_GLD; 1.
CC Oxidoreductase; NAD.
CC NP_BIND 242 247 NAD (ADP PART) (BY SIMILARITY)
CC ACT_SITE 265 265 BY SIMILARITY.
CC ACT_SITE 299 299 BY SIMILARITY.
CC SEQUENCE 497 AA; 53809 MW; 47CA53FF08450F8 CF654;

Query Match 57 18; Score 1455; DR 1; length 497;
Best Local Similarity 56.08; Pred. NO.le-87;
Matches 277; Conservative 83; Mismatches 131. Indels 4; Gaps 3;

```

RESULT 4

DHA6_YEAST

ID DHA6_YEAST

AC P54115; Q02783;

DT 01-OCT-1996 (Rel. 34, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DE Magnesium-activated aldehyde dehydrogenase,

DE (Mg(2+)-activated aldehyde dehydrogenase), cytosolic (EC 1.2.1.3)

DE ALDH OK ALDH1 OR YP0061W OR LPE9.

GN Saccharomyces cerevisiae (Baker's yeast).

OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OX Saccharomycotales; Saccharomycetaceae; Saccharomycetes.

RN NCBI_TaxID=4932;

RN [1]

RP SEQUENCE FROM N A, AND CHAPARTEPKEIZATION.

PC STRAIN=DSBY939;

PX MEDLINE=9813277; PubMed=9473035;

RA Wang X, Mann C J, Bai Y, Ni L, Weiner H.;

FT "Molecular cloning, characterization, and potential roles of cytosolic

FT and mitochondrial aldehyde dehydrogenases in ethanol metabolism in

FT Saccharomyces cerevisiae.";

FL J. Bacteriol. 180:822-830(1998).

RN [2]

RN SEQUENCE FROM N A.

PC STRAIN=S288C / AB972;

PX MEDLINE=97413271; PubMed=9169875;

RA Russry H, Strams F K, Ahmed A, Albrecht K, Allen F, Ansorge W,;

RA Araujo R, Avaricio A, Barrell B G, Badcock K, Benes V,;

RA Botstein D, Bowman S, Bruckner M, Carpenter J, Cherry J M,;

RA Chung E, Churcher C M, Cosser F, Davis K, Davis R W,;

RA Dietrich F S, Delius H, DiPaolo T, Dubois E, Duesterhoeft A,;

RA Duncan M, Floeth M, Fortin N, Friesen J P, Fritz C, Goffeau A,;

RA Hall J, Hebling U, Heumann K, Hilbert H, Hillier L,;

RA Hunnicke-Smith S, Hyman R, Johnston M, Kaiman S, Kleine K,;

RA Komp C, Kurdi O, Lashkari D, Low H, Lin A, Lin D, Louis F J,;

	111 11 111 11 11111111 111 11111111111111111111
Db	184 PALAMGNVCTLKPAAVTFLNALVFASLKKVGTPACVWNIIVPQGRIVTAALINIPPTPK 243
QY	241 VAFITGATGGH- YGSAAAGCEKKEVLEIGDQSPNIFVAFAKFKKAVGINILGIYYNSDE 299
Lb	244 LAFISTEVSSESVAVISSSNRFEETILELFGESABEVFPANFILLINIVNGTFKNAGQ 403
QY	400 VCAGSRVVYPESIYKFIFFPPKAASES-IKVGPDPSTFPGACTSQMINKLIKYYDI 458
Db	304 ICSSGSRIYVOEGIYDELLAAPRAYLETEIKVGNPDKANPQSALTNRQDFDIMNYDI 463
QY	369 GKNGSATLTGGEPLSGSKGFIEKTVGVGVKEMIRLVKEETEDVPVTVIKPSADEVNM 418
Db	364 GKKECAKLTCGEKVCGRGYFTPIYVYGVNEIMRVLKEETEPDVYVIKAKTLPESVEM 423
QY	419 ANDSEYGLAACHTSNTALKVKADKVNACTVYNINOFHVAVTGGEFNASGLAREMSV 478
Lb	424 ANSFPELGSGSLTEPSLSGLKVMMLKGVWNIINYNDISKVPTELAVKSGYREMGTE 483
QY	479 DALQNYLQVRKAVRAKL 494
Db	484 EYVHYAYTEVRKAVRIKL 499

RESULT 5			
ID	DHAL_EMENI	STANDARD:	PRT: 497 AA.
DT	AC P08157;		
DT	01-AUG-1988 (Rel. 08, Created)		
DT	01-AUG-1988 (Rel. 08, Last sequence update)		
DE	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Aldehyde dehydrogenase (EC 1.2.1.3) (AlDH)		
OS	ALDA OR ASPA.		
OS	Emericella nidulans (Aspergillus nidulans).		
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;		
OX	Eurotiales; Trichocomaceae; Emericella.		
RN	NCHI_TaxID=5072;		
RP	[1]		
RP	SEQUENCE FROM N.A.		

MEDLINE=87248080; PubMed=1046652;
 RA Pickett M., Gwynne D.L., Buxton F.P., Elliott R., Davies K.W.,
 Kerkington R.A., Scanzocchio C., Sealy-Lewis H.M.;
 RA "Cloning and characterization of the alda gene of *Aspergillus*
 RT nidulans";
 RT Gene 51:217-226(1987).
 CC -I- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2O) -> an acid + NADH.
 CC -I- PATHWAY: Ethanol utilization; second step.
 CC -I- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. See <http://www.isdb.ch/sequence/sequence.html>

CC		or send an email to licencese@sb.sib.ch).
EMBL:	M16197; AA33293.1;	-
FIR:	A29875; A29055.	-
HSSP:	P05091; ICW3.	-
InterPro:	IPR002086; Aldehyde_dehydr.	-
PFam:	PF00171; aldedeh_1.	-
PROSITE:	PS00070; ALDEHYDE_DEHYDR_CYS; 1.	-
PROSITE:	PS00687; ALDEHYDE_DEHYDR_GLD; 1.	-
KW	oxidoreductase; NAD,	-
NP_RIN%	241	246 NAD (AID PART) (BY SIMILARITY)
ACT_SITE:	264	264 BY SIMILARITY
ACT_SITE:	298	298 BY SIMILARITY
SEQUENCE:	497 AA; 54088 MW; 1711EEAA94SE571B CPO64;	-
Query Match	95.4%; Score 1412.6; IDV 1; Length 497;	-
Best Local Similarity	94.5%; Prod. No. 6, 1 st R _g ;	-
Matches	270; Conservative	142; 100% LS
		Gauss

QY 1 MSILPMTVPDKLENGLEVEOPTGLFINNKFVPSKONKTEFEVINTEFEETCHIEYREDD 60
 DB 1 MSOLFTEITFPV---IKYEQUGUGFINNEFVKGVEGKTFQVINNEKVFITSVHEATEKD 57
 QY 61 VFFAVCAADPAFENFSGWNGTIPIDPKALFYPIAFIPECKIVIVIASIPFUNGKALSSSG 120
 DB 58 VDVAVAAAPAF-FEPWPCVIVSEPGIILNKLAUMERUULFLAAIESLNGKAFIMAKV 116
 QY 121 DVLIVINYLSKSGADKIDGMIDTGTHTPSYTKROPGLVCGGQIIPWNPPLLMWAKIA 180
 DB 117 DLANSIGGLRYAYAGWADRIHQITPILINPEHIIYTHHEFVSVGQIIPWNPPLLMWSKIG 176
 QY 181 PALVTGNTVWLKTAESTPLSALYVSKYIPQAGIPPGVINISVSGKIVVEATINHPKIKK 240
 DB 177 PAVAGNTVWLKTAQTPLSALYAAKLIKEAPFAGVINISVSGRTAGAAIASHMDIDK 236
 QY 241 VAGTGTATGRHIYOSAA-AGLKVVTELGKSPNIPADAEAKKAVONIIILGIYNSGE 299
 DB 237 VAGTGTSTVGPPIIQAASAKSLKVVTELGKSPNIPVDDADIDNAISWANFGIFENHGQ 296
 QY 300 VCCAGSRVYVESIYDKLEEFKKAESIKYGDPEDESTFOGQTSOMOLNKLILKYVDIG 359
 DB 297 CCAGSPILVDEIYDKLEEFKKAESIKYGDPEDESTFOGQTSOMOLNKLILKYVDIG 356
 QY 360 KNEGATLITGGERLGSKGIFYIKPTVFGDKDMRIVKEEIPGPVTVITKFSADDEVINMA 419
 DB 357 KKAGATVATGGDRHNGEYFQPTVTDVSDMKIAQEEIPGPVTVITKFSADDEVINMA 416
 QY 420 NUSEVGLAAGIHTSNINTALKVADRVNAGTVWINTYDFHVAHPFGFNGSGLGKREMSVD 479
 DB 417 NSTDYGAAAVHTKNVTAIRVSNALKAGTVWINTYDFHVAHPFGFNGSGLGKREMSVD 476
 QY 480 ALONYLQVKAERAKL 494
 DB 477 ALENYTOIKTVHYRL 491

RESULT 6
 DHAL_CLAHE STANDARD, PRT, 496 AA
 AC P40108;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Aldehyde dehydrogenase (EC 1.2.1.3) (ALDDH) (Allergen Cla h 3) (Cla h
 III)
 GN CLAH3.
 OS Cladosporium herbarum
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina;
 OC Dothideomycetes et Chaetothyrionomycetes incertae sedis;
 OC Mycosphaerellaceae; mitosporic Mycosphaerellaceae; Cladosporium.
 OX NCBI_TaxID=29918;
 RN [1]
 RP SOURCE FROM N.A.
 RC STRAIN=260202-Berlin;
 RX MEDLINE=95206305; PubMed=7898496;
 RA Achatz G., Oberkofler H., Lechenauer E., Simon B., Unger A.,
 RA Kandlauer P., Ehner C., Prillinger H., Kraft P., Breitenbach M.;
 RT "Molecular cloning of major and minor allergens of Alternaria
 RT alternata and Cladosporium herbarum";
 RL Mol. Immunol. 32:213-227(1995).
 CC -/- CATALYTIC ACTIVITY An aldehyde + NAD(+) -> H(2)O - an acid + NADH.
 CC -/- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -/- SIMILARITY: LONGTON TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 CC -/- CAUTION: Two distinct proteins have been termed allergen
 CC cla h 3
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcements/>
 CC or send an email to info@isb-sib.ch).
 CC -----
 DR EMBL: X78228; CAAS5072.1; -;
 DR HSSP: P05091; ICW3.
 LF Protein: 119002046; Aldehyde_dehydr.
 DR Pfam: PF00171; aldedh; 1.
 DR Pfam: PS0070; ALDEHYDE_DEHYDROGENASE; 1.
 DR PROSITE: PS00047; ALDEHYDE_DEHYDROGENASE; 1.
 KW Oxidoreductase; NAD; Allergen.
 FT ACI_SITE 263 263 BY SIMILARITY.
 FT ACT_SITE 296 296 BY SIMILARITY.
 SQ SEQUENCE 496 AA: 5397 MW: 43376606.167996 GPC64;
 Query Match 54.98; Score 1399; DB 1; Length 496;
 Best Local Similarity 53.98; Pred No. 4 6a-84;
 Matches 263; Conservative 89; Mismatches 145; Indels 2; Gaps 2.
 QY 7 IVPIKLPNGLEYEQVIGLFINNKFVPSKONKTEFEVINTEFEETCHIEYREDDVLEAVQ 66
 DB 3 SVQLETPHSKGYEPTGLFINNEFVKGVEGKTFQVINNEKVFITSVHEATEKDVIAVA 62
 QY 67 AADAFASNGSWSGILPIIDPKALFYPIAFIPECKIVIVIASIPFUNGKALSSSGIIVHVI 126
 DB 63 AAPQAF-FEPWPCVIVSEPGIILNKLAUMERUULFLAAIESLNGKAFIMAKV 121
 QY 127 NYLKSSAFPAKILGKMDITGKTHESVTKRPLGVVQIIPWNPPLLMWAKIAPALVTG 186
 DB 122 GCLRYVGGWADKITGKVIDITPTDFTNVYKKEPIGVCRSDHSLPLLMWAKIAPALVTG 181
 QY 187 NTNVIKTAESTPLSALYVSKYIPQAGIPPGVINISVSGKIVVEATINHPKIKKVAFTGS 246
 DB 182 NTNVLKTAQTPGLGLVAASLVKEAGPPGPPVINISVSGKIVVEATINHPKIKKVAFTGS 241
 QY 247 TATGPHIYOSAA-AGLKVVTELGKSPNIPVADAEAKKAVONIIILGIYNSFVACGS 305
 DB 242 TVVGTILKAAASNLKVVTELGKSPNIPVADAEAKKAVONIIILGIYNSFVACGS 301
 QY 306 RVYVESIYDKLEEFKKAESIKYGDPEDESTFOGQTSOMOLNKLILKYVDIGKNEGAT 365
 DB 302 RVYVESIYDKLEEFKKAESIKYGDPEDESTFOGQTSOMOLNKLILKYVDIGKNEGAT 361
 QY 366 LITGERLGSKGIFYIKPTVFGDKDMRIVKEEIPGPVTVITKFSADDEVINMANSEYG 425
 DB 362 VETGSGRKGSGYF:ETPIFNVTELMKIVKEEIPGVCIAKFK:KEDAIKLSNASIYG 421
 QY 426 LAAGTHTSNINTALKVADRVNAGTVWINTYDFHVAHPFGFNGSGLGKREMSVULQNYL 485
 DB 422 LAAAVTETNLTATIVSNALKACTVWNTYTHIQMPTGYKESG:CRELGEDALANYT 481
 QY 486 QVKAERAKL 494
 DB 482 QTKVTSIRL 490

RESULT 7
 DHAL_ALIAL STANDARD, PRT, 495 AA.
 AC P42041;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 FT 15-JUN-2002 (Rel. 41, Last annotation update)
 LE Aldehyde dehydrogenase (EC 1.2.1.3) (ALDDH) (Allergen Alt a 10) (Alt a
 X).
 GN ALTA10.
 OS Alternaria alternata.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
 OC Pleosporales; Pleosporaceae; mitosporic Pleosporaceae; Alternaria.
 OX NCBI_TaxID=5599;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=08-0203-Berlin;
 RX MEDLINE=95206305; PubMed=7898496;


```

Db 24 TSANVAPN...HQFEVFN...FINNEKFAVSETEIIVNLSGLAVIQVAGREKIVD 74
QY 63 EAVQADRAFSGS-WNIDPDRKALYPIAFLEQKQKVIASIEITLNGKA-ISSRG 120
Db 80 KAVKAAAPAFQIGSPWPMASDQGPILYPIADLTERFTYLALETILNFKPVVISYV 139
QY 121 DVDLVINLKSSAGFAUKIDGKMLDTGKTHESYTKRQPLGVCQGLIPWNPFLMMAWKIA 180
Db 140 DIDMVLKCPYAGWAKYHGTIPIDGDFEFTYTHPEFVGVCGGLIPWNPFLMMAWKJ 199
QY 181 PALVTGNTVVLKTAESTPLSALYSKYIPIQAGIPPGVINIVSGFGKIVVEAITNHPKIKK 240
Db 200 PALATCNVVMKVAEQTTLTALYVANLIKEAGFFGVNIVEGFTAGAAIASHEGVDR 259
QY 241 VAFGTGSTATPHYQSAA--ASLKKVILELGSKSPENIVFAALAEKKAQVNIILGIYNSG 298
Db 260 VAFGTGTEVG-HLIQVAAAGSSNLKRVTLFQCKSPNLSLADMDMAVEQAHEFALFNQG 318
QY 299 EVCCAGSRVYVESIYDKFIEEFKAASESIRKVGDPDESTFQCAOTSQMLNKLKYVDI 358
Db 319 QTCGAGSRFTVGVENVYDEFEKSVAKAKSPVGVNPFSPKTEGQVETLQFKKILGYKS 438
QY 359 GRNEKATLITGGERLGSKGYPIKPIVFGKVKFEMPIVKEFIPGPPVVTITFEKSADEVNM 418
Db 379 GQGEAKILCGGSAALWGYPIQPTVFGSLVAKLWMLAKKEIPGPMVILKFKTIEFVWGP 438
QY 419 ANDSEVGLAAGIHTSNITALKVAIDRVNAGTIVNINYNDEHHAVPFGSPFNASGLGEMSV 478
Db 439 ANISYGLAAAVFTKDLLKANYLSQALQAGIWINCYDVEFAGSPFGYKMSGSGRELGE 498
QY 479 DALQNYLQVKAAPAKL 494
Db 499 YGLQAYTEKVTIVKV 514

```

RESULT 9

```

ID DHAM_RAT STANDARD: PRT: 519 AA.
AC P11884;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Aldehyde dehydrogenase, mitochondrial precursor (BC 1 2 1 *) (ALDH
DE class 2 (ALDH1) (ALDH-E2).
GN ALDH2.
OS Rattus norvegicus (Rat)
OC Eukaryota, Metazoa, Chordata, Craniota, Vertebrata, Euteleostomi;
OC Mammalia, Eutheria, Rodentia; Sciurognathi, Muridae, Murinae, Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=89210865; PubMed=254004;
RA Farres J., Guan K.-L., Weiner H.;
RT "Primary structures of rat and bovine liver mitochondrial aldehyde
RT dehydrogenases deduced from cDNA sequences";
RL Eur. J Biochem 180:67-74(1989)
RN [2]
RP SEQUENCE OF 1-29 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=RR134217; PubMed=342060;
RA Farres J., Guan K.-L., Weiner H.;
RT "Sequence of the signal peptide for rat liver mitochondrial aldehyde
RT dehydrogenase.";
RL Biochem. Biophys. Res. Commun 150:1083-1087(1988)
RN [3]
RP SEQUENCE OF 1-19.
RC TISSUE=Liver;
RX MEDLINE=91378548; PubMed=1898068;
RA Jeng J., Weiner H.;
RT "Purification and characterization of catalytically active precursor
RT of rat liver mitochondrial aldehyde dehydrogenase expressed in

```

```

E1 Escherichia coli."
K1 Arch. Biochem. Biophys. 189:114-122(1981).
CC -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC -!- PATHWAY: Ethanol utilization; second step.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC
CC This SWISS-PROT entry is outdated. It is provided through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
FMRL: X14977; CAA33101.1; -
DF EMBL, M12030, AAA40719.1; -
DR PIR: A27713; A27713.
DR PIR: S03564; S03564.
DR PIR: S17492; S17492.
DR HSP: P05091; LCW3.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldehyd; 1.
DF PFSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DF PFSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD; Acetylation; Mitochondrion; Transit peptide.
FT TRANSIT 1 19 MITOCHONDRION.
FT MOD_RES 20 519 ALDEHYDE DEHYDROGENASE.
FT NP_BIND 20 20 ACETYLATION (PROBABLE).
FT NP_BIND 264 269 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 287 287 BY SIMILARITY.
FT ACT_SITE 321 321 BY SIMILARITY.
SQ SEQUENCE 519 AA; 56488 MW; 75c748202e1333e5 CRC64;
Query Match 50.0%; Score 1274.5; DB 1; Length 519;
Post local Similarity 50.8%; Pred. No. 6,4e-76;
Matches 252; Conservative 85; Mismatches 146; Indels 13; Gaps 6;
QY 7 TVPIKPLNGLEYEQP---TGLFINNKFPVPSKQNTFEVINPSTEEIEICHIEGREDDE 62
Db 24 TSANVAPN---QQVEVPCNQIFINNEHUAIVSKETFTTVPNPTGVCQVAGREKIVD 79
QY 63 EAVQADRAFSGS-WNIDPDRKALYPIAFLEQKQKVIASIEITLNGKA-ISSRG 120
Db 80 KAVKAAAPAFQIGSPWPMASDQGPILYPIADLTERFTYLALETILNFKPVVISYV 139
QY 121 DVDLVINLKSSAGFAUKIDGKMLDTGKTHESYTKRQPLGVCQGLIPWNPFLMMAWKIA 180
Db 140 DIDMVLKCPYAGWAKYHGTIPIDGDFEFTYTHPEFVGVCGGLIPWNPFLMMAWKJ 199
QY 181 PALVTGNTVVLKTAESTPLSALYSKYIPIQAGIPPGVINIVSGFGKIVVEAITNHPKIKK 240
Db 200 PALATCNVVMKVAEQTTLTALYVANLIKEAGFFGVNIVEGFTAGAAIASHEGVDR 259
QY 241 VAFGTGSTATPHYQSAA--ASLKKVILELGSKSPENIVFAALAEKKAQVNIILGIYNSG 298
Db 260 VAFGTGTEVG-HLIQVAAAGSSNLKRVTLFQCKSPNLSLADMDMAVEQAHEFALFNQG 318
QY 299 EVCCAGSRVYVESIYDKFIEEFKAASESIRKVGDPDESTFQCAOTSQMLNKLKYVDI 358
Db 319 QTCGAGSRFTVGVENVYDEFEKSVAKAKSPVGVNPFSPKTEGQVETLQFKKILGYKS 438
QY 359 GRNEKATLITGGERLGSKGYPIKPIVFGKVKFEMPIVKEFIPGPPVVTITFEKSADEVNM 418
Db 379 GQGEAKILCGGSAALWGYPIQPTVFGSLVAKLWMLAKKEIPGPMVILKFKTIEFVWGP 438
QY 419 ANDSEVGLAAGIHTSNITALKVAIDRVNAGTIVNINYNDEHHAVPFGSPFNASGLGEMSV 478
Db 439 ANISYGLAAAVFTKDLLKANYLSQALQAGIWINCYDVEFAGSPFGYKMSGSGRELGE 498
QY 479 DALQNYLQVKAAPAKL 494

```


Proc Natl Acad Sci U S A 82:3771-3775(1985)
[8]
SEQUENCE OF 119-517 FROM N.A.
MEDLINE=85252089; PubMed 4515823.
RA Yoshida A., Ikawa M., Hsu J.C., Tani K :
RT "Molecular abnormality and cDNA cloning of human aldehyde
RT dehydrogenases.";
RL Alcohol 2:103-106(1985).
[9]
PP SEQUENCE OF 214-500 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=87279333; PubMed-3610592;
RA Agarwal D.P., Goede H.W.;
PT "Human aldehyde dehydrogenase isozymes and alcohol sensitivity.";
PL Isozymes Curr. Top. Biol. Med. 16:21-48(1987).
[10]
RN DESCRIPTION OF ORIGIN OF CONFLICTS BETWEEN REF 2 AND DNA SEQUENCES
RX MEDLINE=88005159; PubMed-3653404;
RA Hempel J., Hoeeg J.-O., Joernvall H :
PT "Mitochondrial aldehyde dehydrogenase. Homology of putative targeting
RT sequence to that of carbamyl phosphate synthetase I revealed by
RL correlation of cDNA and protein data.";
REBS Lett. 222:95-98(1987).
[11]
RN VARIANT LYS-504.
RX MEDLINE=84119449; PubMed-6582480;
RA Yoshida A., Huang J.-Y., Ikawa M.;
PT "Molecular abnormality of an inactive aldehyde dehydrogenase variant
RT commonly found in orientals.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:258-261(1984).
[12]
RN VARIANT LYS-496.
RX MEDLINE=96119362; PubMed=8561277;
RA Navrozdovsky A., Tsai S.L., Goldfarb L., Peterson R., Long T.C.,
RA Goldman D.;
PT "Mitochondrial aldehyde dehydrogenase polymorphism in Asian and
RT American Indian populations: detection of new ALDH2 alleles.";
RL Alcohol. Clin. Exp. Res. 19:1105-1110(1995).
[13]
RN X-RAY CRYSTALLOGRAPHY (2.5R ANGSTROMS)
RX MEDLINE=20095857; PubMed-10631996;
RA Ni L., Zhou J., Hurley T.D., Weiner H.;
RT "Human liver mitochondrial aldehyde dehydrogenase: three-dimensional
RT structure and the restoration of solubility and activity of chimeric
RT forms.";
PL Protein Sci. 4:2784-2790(1995)
CC -!- CATALYTIC ACTIVITY. An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC -!- PATHWAY: Ethanol utilization; second step.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -!- DISEASE: A VERY HIGH INCIDENCE OF ACUTE ALCOHOL INTOXICATION IN
CC ORIENTALS AND SOUTH AMERICAN INDIANS, AS COMPARED TO CAUCASIANS,
CC CAN BE DIRECTLY ATTRIBUTED TO AN ENZYMATICALLY IMPAIRED ALDH
CC ISOZYME. THE INACTIVE VARIANT (ALLELE 2 OF ALDH2*2) IS DUE TO
CC SINGLE AMINO ACID EXCHANGE.
CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC -!- CAUTION: THE CONFLICTS BETWEEN THE SEQUENCE DETERMINED IN REF 1
CC AND REF 2 AND THAT IN REF 3 AND 4 ARE PROBABLY ALL DUE TO
CC FRAMESHIFT OR SEQUENCING ERRORS AS DESCRIBED IN REF 5 AND REF. 6;
CC THE SEQUENCE DESCRIBED IN REF 9 DIFFERS FROM THAT SHOWN DUE TO TWO
CC FRAMESHIFTS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M20456; AAA51693.1; -
DR EMBL; M20444; AAA51693.1; JOINED.
DR EMBL; M20445; AAA51693.1; JOINED.

DR EMBL; M20446; AAA51693.1; JOINED.
DR EMBL; M20447; AAA51693.1; JOINED.
DR EMBL; M20448; AAA51693.1; JOINED.
DR EMBL; M20449; AAA51693.1; JOINED.
DR EMBL; M20450; AAA51693.1; JOINED.
DR EMBL; M20451; AAA51693.1; JOINED.
DR EMBL; M20452; AAA51693.1; JOINED.
DR EMBL; M20453; AAA51693.1; JOINED.
DR EMBL; M20454; AAA51693.1; JOINED.
DR EMBL; M20455; AAA51693.1; JOINED.
DR EMBL; M20456; AAA51693.1; JOINED.
DR EMBL; M20457; AAA51693.1; JOINED.
DR EMBL; M20458; AAA51693.1; JOINED.
DR EMBL; M20459; AAA51693.1; JOINED.
DR EMBL; M20460; AAA51693.1; JOINED.
DR EMBL; M20461; AAA51693.1; JOINED.
DR EMBL; M20462; AAA51693.1; JOINED.
DR EMBL; M20463; AAA51693.1; JOINED.
DR EMBL; M20464; AAA51693.1; JOINED.
DR EMBL; M20465; AAA51693.1; JOINED.
DR EMBL; M20466; AAA51693.1; JOINED.
DR EMBL; M20467; AAA51693.1; JOINED.
DR EMBL; M20468; AAA51693.1; JOINED.
DR EMBL; M20469; AAA51693.1; JOINED.
DR EMBL; M20470; AAA51693.1; JOINED.
DR EMBL; M20471; AAA51693.1; JOINED.
DR EMBL; M20472; AAA51693.1; JOINED.
DR EMBL; M20473; AAA51693.1; JOINED.
DR EMBL; M20474; AAA51693.1; JOINED.
DR EMBL; M20475; AAA51693.1; JOINED.
DR EMBL; M20476; AAA51693.1; JOINED.
DR EMBL; M20477; AAA51693.1; JOINED.
DR EMBL; M20478; AAA51693.1; JOINED.
DR EMBL; M20479; AAA51693.1; JOINED.
DR EMBL; M20480; AAA51693.1; JOINED.
DR EMBL; M20481; AAA51693.1; JOINED.
DR EMBL; M20482; AAA51693.1; JOINED.
DR EMBL; M20483; AAA51693.1; JOINED.
DR EMBL; M20484; AAA51693.1; JOINED.
DR EMBL; M20485; AAA51693.1; JOINED.
DR EMBL; M20486; AAA51693.1; JOINED.
DR EMBL; M20487; AAA51693.1; JOINED.
DR EMBL; M20488; AAA51693.1; JOINED.
DR EMBL; M20489; AAA51693.1; JOINED.
DR EMBL; M20490; AAA51693.1; JOINED.
DR EMBL; M20491; AAA51693.1; JOINED.
DR EMBL; M20492; AAA51693.1; JOINED.
DR EMBL; M20493; AAA51693.1; JOINED.
DR EMBL; M20494; AAA51693.1; JOINED.
DR EMBL; M20495; AAA51693.1; JOINED.
DR EMBL; M20496; AAA51693.1; JOINED.
DR EMBL; M20497; AAA51693.1; JOINED.
DR EMBL; M20498; AAA51693.1; JOINED.
DR EMBL; M20499; AAA51693.1; JOINED.
DR EMBL; M20500; AAA51693.1; JOINED.
DR EMBL; M20501; AAA51693.1; JOINED.
DR EMBL; M20502; AAA51693.1; JOINED.
DR EMBL; M20503; AAA51693.1; JOINED.
DR EMBL; M20504; AAA51693.1; JOINED.
DR EMBL; M20505; AAA51693.1; JOINED.
DR EMBL; M20506; AAA51693.1; JOINED.
DR EMBL; M20507; AAA51693.1; JOINED.
DR EMBL; M20508; AAA51693.1; JOINED.
DR EMBL; M20509; AAA51693.1; JOINED.
DR EMBL; M20510; AAA51693.1; JOINED.
DR EMBL; M20511; AAA51693.1; JOINED.
DR EMBL; M20512; AAA51693.1; JOINED.
DR EMBL; M20513; AAA51693.1; JOINED.
DR EMBL; M20514; AAA51693.1; JOINED.
DR EMBL; M20515; AAA51693.1; JOINED.
DR EMBL; M20516; AAA51693.1; JOINED.
DR EMBL; M20517; AAA51693.1; JOINED.
DR EMBL; M20518; AAA51693.1; JOINED.
DR EMBL; M20519; AAA51693.1; JOINED.
DR EMBL; M20520; AAA51693.1; JOINED.
DR EMBL; M20521; AAA51693.1; JOINED.
DR EMBL; M20522; AAA51693.1; JOINED.
DR EMBL; M20523; AAA51693.1; JOINED.
DR EMBL; M20524; AAA51693.1; JOINED.
DR EMBL; M20525; AAA51693.1; JOINED.
DR EMBL; M20526; AAA51693.1; JOINED.
DR EMBL; M20527; AAA51693.1; JOINED.
DR EMBL; M20528; AAA51693.1; JOINED.
DR EMBL; M20529; AAA51693.1; JOINED.
DR EMBL; M20530; AAA51693.1; JOINED.
DR EMBL; M20531; AAA51693.1; JOINED.
DR EMBL; M20532; AAA51693.1; JOINED.
DR EMBL; M20533; AAA51693.1; JOINED.
DR EMBL; M20534; AAA51693.1; JOINED.
DR EMBL; M20535; AAA51693.1; JOINED.
DR EMBL; M20536; AAA51693.1; JOINED.
DR EMBL; M20537; AAA51693.1; JOINED.
DR EMBL; M20538; AAA51693.1; JOINED.
DR EMBL; M20539; AAA51693.1; JOINED.
DR EMBL; M20540; AAA51693.1; JOINED.
DR EMBL; M20541; AAA51693.1; JOINED.
DR EMBL; M20542; AAA51693.1; JOINED.
DR EMBL; M20543; AAA51693.1; JOINED.
DR EMBL; M20544; AAA51693.1; JOINED.
DR EMBL; M20545; AAA51693.1; JOINED.
DR EMBL; M20546; AAA51693.1; JOINED.
DR EMBL; M20547; AAA51693.1; JOINED.
DR EMBL; M20548; AAA51693.1; JOINED.
DR EMBL; M20549; AAA51693.1; JOINED.
DR EMBL; M20550; AAA51693.1; JOINED.
DR EMBL; M20551; AAA51693.1; JOINED.
DR EMBL; M20552; AAA51693.1; JOINED.
DR EMBL; M20553; AAA51693.1; JOINED.
DR EMBL; M20554; AAA51693.1; JOINED.
DR EMBL; M20555; AAA51693.1; JOINED.
DR EMBL; M20556; AAA51693.1; JOINED.
DR EMBL; M20557; AAA51693.1; JOINED.
DR EMBL; M20558; AAA51693.1; JOINED.
DR EMBL; M20559; AAA51693.1; JOINED.
DR EMBL; M20560; AAA51693.1; JOINED.
DR EMBL; M20561; AAA51693.1; JOINED.
DR EMBL; M20562; AAA51693.1; JOINED.
DR EMBL; M20563; AAA51693.1; JOINED.
DR EMBL; M20564; AAA51693.1; JOINED.
DR EMBL; M20565; AAA51693.1; JOINED.
DR EMBL; M20566; AAA51693.1; JOINED.
DR EMBL; M20567; AAA51693.1; JOINED.
DR EMBL; M20568; AAA51693.1; JOINED.
DR EMBL; M20569; AAA51693.1; JOINED.
DR EMBL; M20570; AAA51693.1; JOINED.
DR EMBL; M20571; AAA51693.1; JOINED.
DR EMBL; M20572; AAA51693.1; JOINED.
DR EMBL; M20573; AAA51693.1; JOINED.
DR EMBL; M20574; AAA51693.1; JOINED.
DR EMBL; M20575; AAA51693.1; JOINED.
DR EMBL; M20576; AAA51693.1; JOINED.
DR EMBL; M20577; AAA51693.1; JOINED.
DR EMBL; M20578; AAA51693.1; JOINED.
DR EMBL; M20579; AAA51693.1; JOINED.
DR EMBL; M20580; AAA51693.1; JOINED.
DR EMBL; M20581; AAA51693.1; JOINED.
DR EMBL; M20582; AAA51693.1; JOINED.
DR EMBL; M20583; AAA51693.1; JOINED.
DR EMBL; M20584; AAA51693.1; JOINED.
DR EMBL; M20585; AAA51693.1; JOINED.
DR EMBL; M20586; AAA51693.1; JOINED.
DR EMBL; M20587; AAA51693.1; JOINED.
DR EMBL; M20588; AAA51693.1; JOINED.
DR EMBL; M20589; AAA51693.1; JOINED.
DR EMBL; M20590; AAA51693.1; JOINED.
DR EMBL; M20591; AAA51693.1; JOINED.
DR EMBL; M20592; AAA51693.1; JOINED.
DR EMBL; M20593; AAA51693.1; JOINED.
DR EMBL; M20594; AAA51693.1; JOINED.
DR EMBL; M20595; AAA51693.1; JOINED.
DR EMBL; M20596; AAA51693.1; JOINED.
DR EMBL; M20597; AAA51693.1; JOINED.
DR EMBL; M20598; AAA51693.1; JOINED.
DR EMBL; M20599; AAA51693.1; JOINED.
DR EMBL; M20600; AAA51693.1; JOINED.
DR EMBL; M20601; AAA51693.1; JOINED.
DR EMBL; M20602; AAA51693.1; JOINED.
DR EMBL; M20603; AAA51693.1; JOINED.
DR EMBL; M20604; AAA51693.1; JOINED.
DR EMBL; M20605; AAA51693.1; JOINED.
DR EMBL; M20606; AAA51693.1; JOINED.
DR EMBL; M20607; AAA51693.1; JOINED.
DR EMBL; M20608; AAA51693.1; JOINED.
DR EMBL; M20609; AAA51693.1; JOINED.
DR EMBL; M20610; AAA51693.1; JOINED.
DR EMBL; M20611; AAA51693.1; JOINED.
DR EMBL; M20612; AAA51693.1; JOINED.
DR EMBL; M20613; AAA51693.1; JOINED.
DR EMBL; M20614; AAA51693.1; JOINED.
DR EMBL; M20615; AAA51693.1; JOINED.
DR EMBL; M20616; AAA51693.1; JOINED.
DR EMBL; M20617; AAA51693.1; JOINED.
DR EMBL; M20618; AAA51693.1; JOINED.
DR EMBL; M20619; AAA51693.1; JOINED.
DR EMBL; M20620; AAA51693.1; JOINED.
DR EMBL; M20621; AAA51693.1; JOINED.
DR EMBL; M20622; AAA51693.1; JOINED.
DR EMBL; M20623; AAA51693.1; JOINED.
DR EMBL; M20624; AAA51693.1; JOINED.
DR EMBL; M20625; AAA51693.1; JOINED.
DR EMBL; M20626; AAA51693.1; JOINED.
DR EMBL; M20627; AAA51693.1; JOINED.
DR EMBL; M20628; AAA51693.1; JOINED.
DR EMBL; M20629; AAA51693.1; JOINED.
DR EMBL; M20630; AAA51693.1; JOINED.
DR EMBL; M20631; AAA51693.1; JOINED.
DR EMBL; M20632; AAA51693.1; JOINED.
DR EMBL; M20633; AAA51693.1; JOINED.
DR EMBL; M20634; AAA51693.1; JOINED.
DR EMBL; M20635; AAA51693.1; JOINED.
DR EMBL; M20636; AAA51693.1; JOINED.
DR EMBL; M20637; AAA51693.1; JOINED.
DR EMBL; M20638; AAA51693.1; JOINED.
DR EMBL; M20639; AAA51693.1; JOINED.
DR EMBL; M20640; AAA51693.1; JOINED.
DR EMBL; M20641; AAA51693.1; JOINED.
DR EMBL; M20642; AAA51693.1; JOINED.
DR EMBL; M20643; AAA51693.1; JOINED.
DR EMBL; M20644; AAA51693.1; JOINED.
DR EMBL; M20645; AAA51693.1; JOINED.
DR EMBL; M20646; AAA51693.1; JOINED.
DR EMBL; M20647; AAA51693.1; JOINED.
DR EMBL; M20648; AAA51693.1; JOINED.
DR EMBL; M20649; AAA51693.1; JOINED.
DR EMBL; M20650; AAA51693.1; JOINED.
DR EMBL; M20651; AAA51693.1; JOINED.
DR EMBL; M20652; AAA51693.1; JOINED.
DR EMBL; M20653; AAA51693.1; JOINED.
DR EMBL; M20654; AAA51693.1; JOINED.
DR EMBL; M20655; AAA51693.1; JOINED.
DR EMBL; M20656; AAA51693.1; JOINED.
DR EMBL; M20657; AAA51693.1; JOINED.
DR EMBL; M20658; AAA51693.1; JOINED.
DR EMBL; M20659; AAA51693.1; JOINED.
DR EMBL; M20660; AAA51693.1; JOINED.
DR EMBL; M20661; AAA51693.1; JOINED.
DR EMBL; M20662; AAA51693.1; JOINED.
DR EMBL; M20663; AAA51693.1; JOINED.
DR EMBL; M20664; AAA51693.1; JOINED.
DR EMBL; M20665; AAA51693.1; JOINED.
DR EMBL; M20666; AAA51693.1; JOINED.
DR EMBL; M20667; AAA51693.1; JOINED.
DR EMBL; M20668; AAA51693.1; JOINED.
DR EMBL; M20669; AAA51693.1; JOINED.
DR EMBL; M20670; AAA51693.1; JOINED.
DR EMBL; M20671; AAA51693.1; JOINED.
DR EMBL; M20672; AAA51693.1; JOINED.
DR EMBL; M20673; AAA51693.1; JOINED.
DR EMBL; M20674; AAA51693.1; JOINED.
DR EMBL; M20675; AAA51693.1; JOINED.
DR EMBL; M20676; AAA51693.1; JOINED.
DR EMBL; M20677; AAA51693.1; JOINED.
DR EMBL; M20678; AAA51693.1; JOINED.
DR EMBL; M20679; AAA51693.1; JOINED.
DR EMBL; M20680; AAA51693.1; JOINED.
DR EMBL; M20681; AAA51693.1; JOINED.
DR EMBL; M20682; AAA51693.1; JOINED.
DR EMBL; M20683; AAA51693.1; JOINED.
DR EMBL; M20684; AAA51693.1; JOINED.
DR EMBL; M20685; AAA51693.1; JOINED.
DR EMBL; M20686; AAA51693.1; JOINED.
DR EMBL; M20687; AAA51693.1; JOINED.
DR EMBL; M20688; AAA51693.1; JOINED.
DR EMBL; M20689; AAA51693.1; JOINED.
DR EMBL; M20690; AAA51693.1; JOINED.
DR EMBL; M20691; AAA51693.1; JOINED.
DR EMBL; M20692; AAA51693.1; JOINED.
DR EMBL; M20693; AAA51693.1; JOINED.
DR EMBL; M20694; AAA51693.1; JOINED.
DR EMBL; M20695; AAA51693.1; JOINED.
DR EMBL; M20696; AAA51693.1; JOINED.
DR EMBL; M20697; AAA51693.1; JOINED.
DR EMBL; M20698; AAA51693.1; JOINED.
DR EMBL; M20699; AAA51693.1; JOINED.
DR EMBL; M20700; AAA51693.1; JOINED.
DR EMBL; M20701; AAA51693.1; JOINED.
DR EMBL; M20702; AAA51693.1; JOINED.
DR EMBL; M20703; AAA51693.1; JOINED.
DR EMBL; M20704; AAA51693.1; JOINED.
DR EMBL; M20705; AAA51693.1; JOINED.
DR EMBL; M20706; AAA51693.1; JOINED.
DR EMBL; M20707; AAA51693.1; JOINED.
DR EMBL; M20708; AAA51693.1; JOINED.
DR EMBL; M20709; AAA51693.1; JOINED.
DR EMBL; M20710; AAA51693.1; JOINED.
DR EMBL; M20711; AAA51693.1; JOINED.
DR EMBL; M20712; AAA51693.1; JOINED.
DR EMBL; M20713; AAA51693.1; JOINED.
DR EMBL; M20714; AAA51693.1; JOINED.
DR EMBL; M20715; AAA51693.1; JOINED.
DR EMBL; M20716; AAA51693.1; JOINED.
DR EMBL; M20717; AAA51693.1; JOINED.
DR EMBL; M20718; AAA51693.1; JOINED.
DR EMBL; M20719; AAA51693.1; JOINED.
DR EMBL; M20720; AAA51693.1; JOINED.
DR EMBL; M20721; AAA51693.1; JOINED.
DR EMBL; M20722; AAA51693.1; JOINED.
DR EMBL; M20723; AAA51693.1; JOINED.
DR EMBL; M20724; AAA51693.1; JOINED.
DR EMBL; M20725; AAA51693.1; JOINED.
DR EMBL; M20726; AAA51693.1; JOINED.
DR EMBL; M20727; AAA51693.1; JOINED.
DR EMBL; M20728; AAA51693.1; JOINED.
DR EMBL; M20729; AAA51693.1; JOINED.
DR EMBL; M20730; AAA51693.1; JOINED.
DR EMBL; M20731; AAA51693.1; JOINED.
DR EMBL; M20732; AAA51693.1; JOINED.
DR EMBL; M20733; AAA51693.1; JOINED.
DR EMBL; M20734; AAA51693.1; JOINED.
DR EMBL; M20735; AAA51693.1; JOINED.
DR EMBL; M20736; AAA51693.1; JOINED.
DR EMBL; M20737; AAA51693.1; JOINED.
DR EMBL; M20738; AAA51693.1; JOINED.
DR EMBL; M20739; AAA51693.1; JOINED.
DR EMBL; M20740; AAA51693.1; JOINED.
DR EMBL; M20741; AAA51693.1; JOINED.
DR EMBL; M20742; AAA51693.1; JOINED.
DR EMBL; M20743; AAA51693.1; JOINED.
DR EMBL; M20744; AAA51693.1; JOINED.
DR EMBL; M20745; AAA51693.1; JOINED.
DR EMBL; M20746; AAA51693.1; JOINED.
DR EMBL; M20747; AAA51693.1; JOINED.
DR EMBL; M20748; AAA51693.1; JOINED.
DR EMBL; M20749; AAA51693.1; JOINED.
DR EMBL; M20750; AAA51693.1; JOINED.
DR EMBL; M20751; AAA51693.1; JOINED.
DR EMBL; M20752; AAA51693.1; JOINED.
DR EMBL; M20753; AAA51693.1; JOINED.
DR EMBL; M20754; AAA51693.1; JOINED.
DR EMBL; M20755; AAA51693.1; JOINED.
DR EMBL; M20756; AAA51693.1; JOINED.
DR EMBL; M20757; AAA51693.1; JOINED.
DR EMBL; M20758; AAA51693.1; JOINED.
DR EMBL; M20759; AAA51693.1; JOINED.
DR EMBL; M20760; AAA51693.1; JOINED.
DR EMBL; M20761; AAA51693.1; JOINED.
DR EMBL; M20762; AAA51693.1; JOINED.
DR EMBL; M20763; AAA51693.1; JOINED.
DR EMBL; M20764; AAA51693.1; JOINED.
DR EMBL; M20765; AAA51693.1; JOINED.
DR EMBL; M20766; AAA51693.1; JOINED.
DR EMBL; M20767; AAA51693.1; JOINED.
DR EMBL; M20768; AAA51693.1; JOINED.
DR EMBL; M20769; AAA51693.1; JOINED.
DR EMBL; M20770; AAA51693.1; JOINED.
DR EMBL; M20771; AAA51693.1; JOINED.
DR EMBL; M20772; AAA51693.1; JOINED.
DR EMBL; M20773; AAA51693.1; JOINED.
DR EMBL; M20774; AAA51693.1; JOINED.
DR EMBL; M20775; AAA51693.1; JOINED.
DR EMBL; M20776; AAA51693.1; JOINED.
DR EMBL; M20777; AAA51693.1; JOINED.
DR EMBL; M20778; AAA51693.1; JOINED.
DR EMBL; M20779; AAA51693.1; JOINED.
DR EMBL; M20780; AAA51693.1; JOINED.
DR EMBL; M20781; AAA51693.1; JOINED.
DR EMBL; M20782; AAA51693.1; JOINED.
DR EMBL; M20783; AAA51693.1; JOINED.
DR EMBL; M20784; AAA51693.1; JOINED.
DR EMBL; M20785; AAA51693.1; JOINED.
DR EMBL; M20786; AAA51693.1; JOINED.
DR EMBL; M20787; AAA51693.1; JOINED.
DR EMBL; M20788; AAA51693.1; JOINED.
DR EMBL; M20789; AAA51693.1; JOINED.
DR EMBL; M20790; AAA51693.1; JOINED.
DR EMBL; M20791; AAA51693.1; JOINED.
DR EMBL; M20792; AAA51693.1; JOINED.
DR EMBL; M20793; AAA51693.1; JOINED.
DR EMBL; M20794; AAA51693.1; JOINED.
DR EMBL; M20795; AAA51693.1; JOINED.
DR EMBL; M20796; AAA51693.1; JOINED.
DR EMBL; M20797; AAA51693.1; JOINED.
DR EMBL; M20798; AAA51693.1; JOINED.
DR EMBL; M20799; AAA51693.1; JOINED.
DR EMBL; M20800; AAA51693.1; JOINED.
DR EMBL; M20801; AAA51693.1; JOINED.
DR EMBL; M20802; AAA51693.1; JOINED.
DR EMBL; M20803; AAA51693.1; JOINED.
DR EMBL; M20804; AAA51693.1; JOINED.
DR EMBL; M20805; AAA51693.1; JOINED.
DR EMBL; M20806; AAA51693.1; JOINED.
DR EMBL; M20807; AAA51693.1; JOINED.
DR EMBL; M20808; AAA51693.1; JOINED.
DR EMBL; M20809; AAA51693.1; JOINED.
DR EMBL; M20810; AAA51693.1; JOINED.
DR EMBL; M20811; AAA51693.1; JOINED.
DR EMBL; M20812; AAA51693.1; JOINED.
DR EMBL; M20813; AAA51693.1; JOINED.
DR EMBL; M20814; AAA51693.1; JOINED.
DR EMBL; M20815; AAA51693.1; JOINED.
DR EMBL; M20816; AAA51693.1; JOINED.
DR EMBL; M20817; AAA51693.1; JOINED.
DR EMBL; M20818; AAA51693.1; JOINED.
DR EMBL; M20819; AAA51693.1; JOINED.
DR EMBL; M20820; AAA51693.1; JOINED.
DR EMBL; M20821; AAA51693.1; JOINED.
DR EMBL; M20822; AAA51693.1; JOINED.
DR EMBL; M20823; AAA51693.1; JOINED.
DR EMBL; M20824; AAA51693.1; JOINED.
DR EMBL; M20825; AAA51693.1; JOINED.
DR EMBL; M20826; AAA51693.1; JOINED.
DR EMBL; M20827; AAA51693.1; JOINED.
DR EMBL; M20828; AAA51693.1; JOINED.
DR EMBL; M20829; AAA51693.1; JOINED.
DR EMBL; M20830; AAA51693.1; JOINED.
DR EMBL; M20831; AAA51693.1; JOINED.
DR EMBL; M20832; AAA51693.1; JOINED.
DR EMBL; M20833; AAA51693.1; JOINED.
DR EMBL; M20834; AAA51693.1; JOINED.
DR EMBL; M20835; AAA51693.1; JOINED.
DR EMBL; M20836; AAA51693.1; JOINED.
DR EMBL; M20837; AAA51693.1; JOINED.
DR EMBL; M20838; AAA51693.1; JOINED.
DR EMBL; M20839; AAA51693.1; JOINED.
DR EMBL; M20840; AAA51693.1; JOINED.
DR EMBL; M20841; AAA51693.1; JOINED.
DR EMBL; M20842; AAA51693.1; JOINED.
DR EMBL; M20843; AAA51693.1; JOINED.
DR EMBL; M20844; AAA51693.1; JOINED.
DR EMBL; M20845; AAA51693.1; JOINED.
DR EMBL; M20846; AAA51693.1; JOINED.
DR EMBL; M20847; AAA51693.1; JOINED.
DR EMBL; M20848; AAA51693.1; JOINED.
DR EMBL; M20849; AAA51693.1; JOINED.
DR EMBL; M20850; AAA51693.1; JOINED.
DR EMBL; M20851; AAA51693.1; JOINED.
DR EMBL; M20852; AAA51693.1; JOINED.
DR EMBL; M20853; AAA51693.1; JOINED.
DR EMBL; M20854; AAA51693.1; JOINED.
DR EMBL; M20855; AAA51693.1; JOINED.
DR EMBL; M20856; AAA51693.1; JOINED.
DR EMBL; M20857; AAA51693.1; JOINED.
DR EMBL; M20858; AAA51693.1; JOINED.
DR EMBL; M20859; AAA51693.1; JOINED.
DR EMBL; M20860; AAA51693.1; JOINED.
DR EMBL; M20861; AAA51693.1; JOINED.
DR EMBL; M20862; AAA

Query Match		56.23, Score 1433.5, DB 3, Length 497;
Best Local Similarity		55.26, Pred. No. 3.4e 89;
Matches		273, Conservative 28, Mismatches 129, Indels 5, Gaps 3;
QY	1	MSHLPMTVPKLPNGLEYEQPTGLFINNKEVPSPKQNKTFEVIINPSTFEETICHIEGREDU 60
DB	1	MSDLFTTLETPV---IKYEQPLGLFINNEFPKVGESKTFQVINPNSKEKVTISVHEATEKD 57
QY	61	VFAVAAALVAFVAFSGNSWNGIIPIDPKALYFLAELIEUKIVIASIETIUNGKAISSSW 120
DB	58	VDAVAAARAARF-BGPPQVPTSPERGILINKLADLMERDITDIAAFSLUNGKAFPTMAKV 116
QY	121	DVDLVINLYKSSAGFAKIDGPMIDTGTTHFTSTKQPLAVGCGQIIPWNPPLLMWAKTIA 180
DB	117	ULANSIGLRYAAAWAUKIRGQILIDNPETIIVYTRHEPVGCGQIIPWNPPLLMWAKTIG 176
QY	181	PALVTNTVLTKAESTPLSALYYSKYIPOAGIPDPGVINIVSGPQIIPWNPPLLMWAKTIA 240
DB	177	FVAAAGNTVLKIAEQTLKAYVAANLIKKEAGFPGVVNVISGFGKAVAGAAASHMDIDUKVAF 246
QY	241	VAGTGTATGRHIVQSAA-AGLKKVTLLEGGKSPNIVFAADAEELKKAQVNIILQIYYNSGECC 299
DB	237	VAFGSGTGLVGLKIAEQTLKAYVAANLIKKEAGFPGVVNVISGFGKAVAGAAASHMDIDUKVAF 296
QY	300	VCAESPVYVEESYDFIEEFKAASESIKVGDPFDESFGSAGTSMQJNKILKYVDIG 359
DB	297	QVCAESPVYVEESYDFIEEFKAASESIKVGDPFDESFGSAGTSMQJNKILKYVDIG 356
QY	360	KNEGATLITGCEPLGSGYFIKTFIVGDKVEKMPVYKEEIPCTVVTIVTFYKSAFVIRMA 419
DB	357	KKATATVATGSGPHEGNEFYIQTIVTIVTSUKIACERTFGPQVVTIOKPKFAFAIKIG 416
QY	420	NUSEYGLAAGTHINIAIKLVAKVAGIWNINYNVPHRAVFFGEGNASSGLGPRMSVD 479
DB	417	NSHYGLAAAVHTKANVTAIPVSNALIKAGIWINNINYNMISYAPPFGFGSGELGRELGSY 476
QY	480	ALONYQVAVRAKL 494
DB	477	ALENTQIKTVHYRL 491
RESULT 5		
Q9C106		
ID	Q9C106	PRELIMINARY; PRT; 497 AA.
AC	Q9C106;	
DT	01-JUN-2001 (TREMREL. 17, Created)	
DT	01-JUN-2001 (TREMREL. 17, Last sequence update)	
DT	01-MAR-2002 (TREMREL. 20, Last annotation update)	
DE	Aldehyde dehydrogenase ALDH5.	
GN	ALDA.	
OS	Emmericella nidulans (Aspergillus nidulans).	
OC	Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;	
OC	Eurotiales; Trichocomaceae; Emmericella.	
OX	NCBI_TaxID=5072;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE 2126214, PubMed 11102432,	
RA	Filippi M., Marfisi M., Cirpus I., Papanicolaou C., Pelentok R.	
RT	"Regulation of the aldehyde dehydrogenase gene (aldA) and its role in	
RT	the control of the co-inducer level necessary for induction of the	
RT	ethanol utilization pathway in Aspergillus nidulans."	
PL	J Biol Chem 276:6450-6458(2001).	
CC	-I- SIMILARITY: RELINGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.	
DR	PMR1; AF260124; AAK18073.1;	
DR	HSSP; P05091; 1CW3.	
DR	InterPro: IPR002086; Aldehyde_dehydr.	
DR	Pfam: PF00171; aldedh; 1.	
DR	PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.	
DR	PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.	
KW	Oxidoreductase.	
FT	VARIANT 286	191; E > Q.
FT	VARIANT 286	208; G > P.
FT	VARIANT 286	247; R > P.
RESULT 4		
Q9C105		
ID	Q9C105	PRELIMINARY; PRT; 497 AA.
AC	Q9C105;	
DT	01-JUN-2001 (TREMREL. 17, Created)	
DT	01-JUN-2001 (TREMREL. 17, Last sequence update)	
DT	01-MAR-2002 (TREMREL. 20, Last annotation update)	
DE	Aldehyde dehydrogenase ALDH5.	
GN	ALDA.	
OS	Emmericella nidulans (Aspergillus nidulans).	
OC	Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;	
OC	Eurotiales; Trichocomaceae; Emmericella.	
OX	NCBI_TaxID=5072;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE 2126214, PubMed 11102432,	
RA	Filippi M., Marfisi M., Cirpus I., Papanicolaou C., Pelentok R.	
RT	"Regulation of the aldehyde dehydrogenase gene (aldA) and its role in	
RT	the control of the co-inducer level necessary for induction of the	
RT	ethanol utilization pathway in Aspergillus nidulans."	
PL	J Biol Chem 276:6450-6458(2001).	
CC	-I- SIMILARITY: RELINGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.	
DR	PMR1; AF260124; AAK18073.1;	
DR	HSSP; P05091; 1CW3.	
DR	InterPro: IPR002086; Aldehyde_dehydr.	
DR	Pfam: PF00171; aldedh; 1.	
DR	PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.	
DR	PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.	
KW	Oxidoreductase.	
FT	VARIANT 286	191; E > Q.
FT	VARIANT 286	208; G > P.
FT	VARIANT 286	247; R > P.

```

FT VARIANT 409 409 E -> V
SQ SEQUENCE 497 AA: 54138 MW: 9C6CP62731A25BD CRC64:

Query Match
Best Local Similarity 56.1%; Score 1429.5; DB 3; Length 497,
Matches 273; Conservative 87; Mismatches 130; Indels 5; Gaps 3;

QY 1 MSHLPMTVPVLIKIPNGLEYEQPTCLFINNKFVPSKONKTFEVIINPSTEEIECHITVGRERD 60
DB 1 MSDFLTITETPV IKYEUGLGFINNEFVGVEKTFQVFNPSNEKVIITSVHEAIEKD 57
QY 61 VEFVQAADRAFSNGSWNGIDPIDRGKALYRLAELEGKULVIASITELONGKALISSSKG 120
DB 58 VDVAVAARAARAF-EGPWRQVTPSEKGIINKLADLMERDIDFLAAIESLNDKCAFIMAKV 116
QY 121 DVDLVINYLKSSAGFAIKRGMIIIGPITHFSYTKRQPLGVVAGQIIPWNPFLIMMAWKIA 180
DB 117 DLANSIGCLRYAGWADKIHGQTIIDNPETLITRHEPVGVGSGQIIPWNPFLIMMSWKIG 176
QY 181 PALVTGNTVVKTAESTPLSALYVSKYIPQAGIPPGVINIVSGEKLIVVEALTNHPKIKK 240
DB 177 PAVAAGNTVVKTAETPTLSALYAAKLIKFAFPACVINIVSGRGTAGAAISSHMDIDK 236
QY 241 VAFGTSTATGRHIYQSAA-AGLKVKVTLRIGKSPNIVFADAEIKKAVGNIIILGIYNSGE 299
DB 237 VAFGTSTLVGRTILQAAKSNLKKVTLLEGKSPNIVFDDADIDNAISWANFGIFENHGQ 296
QY 300 VCCAGSRVYVEESIDKFEETEEKAASESILKVGDPDESTEGAGTSQMLNKLKLYVDIG 359
DB 297 CCAGSRPLVQEGIYDKFVARKERAKKNKGNPFQDTFQGVQSLOQDPTIMFYNHG 356
QY 360 KNEGATLITGGERLGSKGFIKPTVFGDKEDMRIVKEELFGPVVTVTKFKSADEVINMA 419
DB 357 KKAGATVATGDRHNEGFIQPTVFTDVTSDMKTAQEFIFGPPVTVTKFKDEAEAIKIG 416
QY 420 NDSFYGLAAGIHTSNINTALKVADRVNAGTVINYNDFHRAVPPGPGNASCLPFMSVD 479
DB 417 NSTDYGILAAAVHTKNVNTAIRVSNALKAGTVINNNMISYQAPGPGKQSGLGRFGSY 476
QY 480 ALQNYLQVKAVRAKL 494
DB 477 ALENYTQIKTVHYRL 491

RESULT 6
Q9C1Q4 PRELIMINARY; PRT: 497 AA.
AC Q9C1Q4;
DT 01-JUN-2001 (TREMblrel, 17, Created)
DT 01-JUN-2001 (TREMblrel, 17, last sequence update)
DT 01-MAR-2002 (TREMblrel, 20, last annotation update)
DE Aldehyde dehydrogenase ALDH57.
GN ALDA.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes.
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21269214; PubMed=11102439;
RA Flippin M., Mathion M., Girpus I., Panov G., Felsenberg B.;
RT "Regulation of the aldehyde dehydrogenase gene (aldA) and its role in
RT the control of the co-inducer level necessary for induction of the
RT ethanol utilization pathway in Aspergillus nidulans.";
RL J. Biol. Chem. 276:6950-6958(2001).
CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
DR EMBL: AF260125; AAK18074.1;
DR HSSP: P05091; 1CW3.
DR InterPro: IPR002086; Aldehyde-dehydr.
DR Pfam: PF00171; aldedh; 1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW oxidoreductase.

FT VARIANT 338 338 S -> G
SQ SEQUENCE 497 AA: 54168 MW: 54F425A84B406475 CRC64:

Query Match
Best Local Similarity 54.9%; Score 1424.9; DB 3; Length 497,
Matches 252; Conservative 87; Mismatches 141; Indels 5; Gaps 3;

QY 1 MSHLPMTVPVLIKIPNGLEYEQPTCLFINNKFVPSKONKTFEVIINPSTEEIECHITVGRERD 60
DB 1 MSDFLTITETPV IKYEUGLGFINNEFVGVEKTFQVFNPSNEKVIITSVHEAIEKD 57
QY 61 VEFVQAADRAFSNGSWNGIDPIDRGKALYRLAELEGKULVIASITELONGKALISSSKG 120
DB 58 VDVAVAARAARAF-EGPWRQVTPSEKGIINKLADLMERDIDFLAAIESLNDKCAFIMAKV 116
QY 121 DVDLVINYLKSSAGFAIKRGMIIIGPITHFSYTKRQPLGVVAGQIIPWNPFLIMMAWKIA 180
DB 117 DLANSIGCLRYAGWADKIHGQTIIDNPETLITRHEPVGVGSGQIIPWNPFLIMMSWKIG 176
QY 181 PALVTGNTVVKTAESTPLSALYVSKYIPQAGIPPGVINIVSGEKLIVVEALTNHPKIKK 240
DB 177 PAVAAGNTVVKTAETPTLSALYAAKLIKFAFPACVINIVSGRGTAGAAISSHMDIDK 236
QY 241 VAFGTSTATGRHIYQSAA-AGLKVKVTLRIGKSPNIVFADAEIKKAVGNIIILGIYNSGE 299
DB 237 VAFGTSTLVGRTILQAAKSNLKKVTLLEGKSPNIVFDDADIDNAISWANFGIFENHGQ 296
QY 300 VCCAGSRVYVEESIDKFEETEEKAASESILKVGDPDESTEGAGTSQMLNKLKLYVDIG 359
DB 297 CCAGSRPLVQEGIYDKFVARKERAKKNKGNPFQDTFQGVQSLOQDPTIMFYNHG 356
QY 360 KNEGATLITGGERLGSKGFIKPTVFGDKEDMRIVKEELFGPVVTVTKFKSADEVINMA 419
DB 357 KKAGATVATGDRHNEGFIQPTVFTDVTSDMKTAQEFIFGPPVTVTKFKDEAEAIKIG 416
QY 420 NDSFYGLAAGIHTSNINTALKVADRVNAGTVINYNDFHRAVPPGPGNASCLPFMSVD 479
DB 417 NSTDYGILAAAVHTKNVNTAIRVSNALKAGTVINNNMISYQAPGPGKQSGLGRFGSY 476
QY 480 ALQNYLQVKAVRAKL 494
DB 477 ALENYTQIKTVHYRL 491

RESULT 7
Q14293 PRELIMINARY; PRT: 504 AA.
AC Q14293;
DT 01-NOV-1998 (TREMblrel, 08, Created)
DT 01-NOV-1998 (TREMblrel, 08, last sequence update)
DT 01-JUN-2001 (TREMblrel, 17, last annotation update)
DE Probable aldehyde dehydrogenase (EC 1.2.1.4) (ALDH6).
GN SPAC9E9.09C.
OS Schizosaccharomyces pombe (Pission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes.
OC Schizosaccharomycetales; Schizosaccharomycetaceae.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX SPEAN 472;
RA McDougal R., Bartell R.G., Kallandrom M.A., Wood V.;
RT Submitted (SEP-1997) to the EMBL/Genbank/Tran databases.
CC -!- CATALYTIC ACTIVITY: ALDEHYDE + NAD(+) + H(2O) -> ALD + NADH.
CC -!- PATHWAY: SECOND STEP IN ETHANOL UTILIZATION.
CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
DR EMBL: Z99452; CAB16407.1;
DR HSSP: P05091; 1CW4.
DR InterPro: IPR002086; Aldehyde-dehydr.
DR Pfam: PF00171; aldedh; 1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW oxidoreductase; NAD.

```

[illegible]


```

Db 299 VKKVTLELGGKSPNIVFAIDADLEAVRQSHALFFNOGCCSAGSKTFVEGKIYDEFVAK 358
QY 321 FKAASESIKWDDEPDESTFOCAOTSOMQLNKILKYVIGCKNEAAILITGSEBLGSKGYF1 380
Db 359 AKELIVKTVIGDPFDENTTGPQIDQSVETIMKYIESGKKEGAQLVTGKVKHGGQGYFV 418
QY 381 KPTVFGDVKKEIMRIVKBEIFGPVVTVTKEKSADEVINMANDEYGLAAGIHTSINTALK 440
Db 419 KPTIFANVNDOMKTAQBEIFGPVVIIVRSDSMEELIEKANNTIYGLAAGVVTNDLNLKALQ 478
QY 441 VADRVNAGTVINTYNDHFHIAVDFGGFNAGSLGKREMSVIALONYLOVRAVRAKL 494
Db 479 VANTIRAGSVVNCYDVFDAAPPFGGKQSGTGRGELGYGLAAYTEVKTVTIKV 532

RESULT 13
Q9BV45 PRELIMINARY; PRT: 517 AA.
ID Q9BV45 AC Q9BV45;
DT 01-JUN-2001 (Tremblrel, 17, Created)
DT 01-JUN-2001 (Tremblrel, 17, Last sequence update)
DT 01-MAR-2002 (Tremblrel, 20, Last annotation update)
DE Similar to aldehyde dehydrogenase 5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 1;
RP SEQUENCE FROM N.A.
RC TISSUE=EYE;
RA Strausberg R.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
DF EMBL: AC001619; AAH01619.1; -
DR HSSP: P05091; ICW3
DR InterPro: IPR02086; Aldehyde_dehydr.
DR Pfam: PF00171; aldehyde_1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase.
SQ SEQUENCE 517 AA; 57238 MW; A628E448D1E8689D CRC64;

```

```

Query Match 49.0%; Score 1249.5; DB 4; Length 517;
Best Local Similarity 50.8%; Pred. NO. 1.2e-76;
Matches 248; Conservative 85; Mismatches 148; Indels 9; Gaps 6;
QY 9 PIKPLNGLEYEQPTGLFINNKFVPSKQKTFEVIINPSTEEETCHIEGREDDVEAVQAA 68
Db 28 PILNPD-IPYQ...LFINNQDAVSKKTFPTVNPITGTEVIGHVAESDRAVDPAVKA 83
QY 69 DRAFSNKS-WNGIDPIDRGKALYRLAELEQDKVIASIFTLNKGATSSSPG-DVRLVI 126
Db 84 REAFPGSPMPMPDASPGPLNLNLADIVERPQVYIASIFTLN-KPFESEYALMLDFV; 143
QY 127 NYLKSSGAFADKIDGRMDIDGRHESYTKRQPLQVCGQIDPNPFLMLMANKIALALVTG 186
Db 144 KVRVYTAGWADKWHGKTIPMGQHPFTTHPEVGVGQDILPNPFLVQCKKLALALATG 203
QY 187 NTVVLKTAESTPLSALIVSKYIPQAGIPPGVINTIVSRPKIVVEATINHPKIKVAFTGS 246
Db 204 NTVMKVAEOTPLSALYLASLKEAGFPFGVNNITGVTGPTAAGATAGHMTVTKVAFTGS 263
QY 247 TATGRHIYQSA--AGLKKVTLIELGKSPNIVFAIAELKKAQVONILGLIYNS-EEVYAG 304
Db 264 TEWG-HLLOKAGGSNKRVTILEGKSPSIVLAIDMEHAEVQHEAIFNMNQVYAG 322
QY 305 SRYVVEESIVDKFIEFKAAASESIKVGDPFDESTFOGAQTSOMQLNKILKYVDLGNKAGA 364
Db 323 SPTEVESIVNELEPTVEKAKQKVGNPFEITDQSGPQVQKESPEFVGLVGLQGRECA 382
QY 365 TLTGGERLGSKGYFIKPTVFGQVDFMDPTVKKFETPRGVVITVFKPSADRVNNMSPY 424
Db 383 KLGGGPFGEERGFETKPTVFGVQDDMKIAKEETFGVQPLFKFKKTEEVVERANNTRY 442

```

```

QY 425 GLAAGIHTSINTALKVAURVNACTVWINTYNDHFHIAVDFGGFNAGSLGKREMSVIALONY 484
Db 443 GLAAAVTEHIDFAMFTLALAGAGVWVNYNIV; HTLFGPKESGN-PEFGEDHAKY 502
QY 485 LOVKAVRAKL 494
Db 503 TEVKTVTIKV 512

RESULT 14
Q9FRX7 PRELIMINARY; PRT: 549 AA.
ID Q9FRX7 AC Q9FRX7;
DT 01-MAR-2001 (Tremblrel, 16, Created)
DT 01-MAR-2001 (Tremblrel, 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel, 21, Last annotation update)
DE Aldehyde dehydrogenase ALDH2B.
GN ALDH2B.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzoideae; Oryza.
OX NCBI_TaxID=4530;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=CV, NIPPONBARE; TISSUE LEAVES OF SEEDLINGS;
RA Nakazono M.; Tsuji H.;
KT "Rice mitochondrial aldehyde dehydrogenase ALDH2B."
CC -!- SIMILARITY: BELONGS TO THE EMBL/GenBank/DDBJ databases.
DF EMBL: AB044537; BAB19052.1; -
DR HSSP: P05091; ICW3
DR InterPro: IPR062386; Aldehyde_dehydr.
DR Pfam: PF00171; aldehyde_1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase.
SQ SEQUENCE 549 AA; 59306 MW; 2B3FC2B841FB6BDD CRC64;

```

```

Query Match 48.9%; Score 1247.5; DB 10; Length 549;
Best Local Similarity 49.8%; Pred. NO. 1.8e-76;
Matches 241; Conservative 85; Mismatches 154; Indels 5; Gaps 3;
QY 9 PIKPLNGLEYEQPTGLFINNKFVPSKQKTFEVIINPSTEEETCHIEGREDDVEAVQAA 68
Db 59 PISPPVQVNTQ---LLIDGRVDSAGSKTFTPTDPTGELIAHVAEHAEDINKAVHAA 115
QY 69 DRAFSNKS-WNGIDPIDRGKALYRLAELEQDKVIASIFTLNKGATSSSPG-DVRLVI 127
Db 116 PFAFTECPMPMTAVERSPIELPFAQLTEPHRTEAALETWIKVYAAANIEVIMVAP 176
QY 126 YFSSA-3ALKIDPMDPDE-6-00-3YFPQDQVQGTQVWNPDEIMWAW-LALAVLN 199
Db 176 LMRYYAGWALKDGLAVVIALGPHVQVLIHETGVA-QITPNPFLIMAKVGTALAVGN 235
QY 188 TVVLKTAESTPLSALIVSKYIPQAGIPPGVINTIVSRPKIVVEATINHPKIKVAFTGS 247
Db 236 IYVLTAEVTPLSALFASKLLEAGLPGVNVVSGPGIACAAALASHMWIKLAFIUST 295
QY 248 ATSPHYYSAA-AATPKKVTILEGKSPNIVFAIAELKKAQVONILGLIYNS-EEVYAG 306
Db 296 DTGKVVLELAANSNLSVILELGKSPILMDAIDVHIAVLAHIALEFPNQGGVANSR 355
QY 307 VYVESLYDKFIEFKAAASESIKVGDPFDESTFOGAQTSOMQLNKILKYVDLGNKAGA 366
Db 356 IEVHEKLYDEEVEKAKALAKAVVGDPEKNVQVQVLDDEQENKILRYIKAVLISGANL 415
QY 367 TETGPELSEKGYFIKPTVFGQVDFMDPTVKKFETPRGVVITVFKPSADRVNNMSPY 426
Db 416 VTGVPIDRGYITQPTTESVQNNRPAQEPFGQVSDITFENQNFVTFKANAAGYGI 475
QY 427 AAGIHTSINTALKVAURVNACTVWINTYNDHFHIAVDFGGFNAGSLGKREMSVIALONY 486

```

Db 476 AAGVTNNINTANTILTRALRVGTVVWNCFFVFAAIFPGDYKSSGSGSEKSIUSLKNYIQ 545
QY 487 VKAV 490
Db 536 VKAV 539

RESULT 15

Q9LLR2 PRELIMINARY: PRT; 549 AA.
AC Q9LLR2;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Aldehyde dehydrogenase.
GN ALDH.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Xu B., Xing Y., Zhang J., Zhu S., Hong M.,
RT "Cloning, characterization of cDNA encoding rice aldehyde
dehydrogenase, and analysis of its expression in male-sterile lines";
RL Chih We Sheng, Li Hsieh-Pao 26:206-212(2000).
CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
DR EMBL: AF12665; AAF73828.1; -.
DR HSSP: P05091; ICW3.
DR InterPro: IPR002086; Aldehyde_dehyd.
DR Pfam: PF00171; aldedh.1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS.1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU.1.
KW Oxidoreductase.
SQ SEQUENCE 549 AA: 59320 MW: 8F0CC6449F34RFR2 CFC64;

Query Match 48 qk; Score 1246.5; DB 10; Length 549;
Best Local Similarity 49.8%; Pred. No. 2.1e-76;
Matches 241; Conservative 87; Mismatches 151; Indels 5; Gaps 3;
QY 9 FIKLNGLEYEQTGLFNNKVPKSKNTFEVINPSTEEELCHIEYEGRELDVVEEAQAA 68
DB 59 PTSPFVQVNYTQ--LITGKREVLSASGKTPPTLCPVIGELLIAHVAEGDAEDINPAVHAA 115
QY 69 PRAFSGQWNIIDPIORAKALYRIAELETQDKQVIAETLEDNCKAIS-SSRGDQDLVIN 127
DB 116 RFAFDGPPWPKMTAYFRSPILLRFDADLIEKHNDIEALETWONGKPYAQAAQTEVPWVAR 175
QY 128 YLKSSAGFADKIDGRMIDTGRTHFSYTKRQPLGVCGQIIPWNPFLMWMKIAAPALVTGN 187
DB 176 LMKYVAGWALKLHGLVWFADGPHVQVLHIEPIGVAGQIIPWNPFLMFAWKVGPALACGN 235
QY 188 TVVLKTAESTPLSALYSKYIPOAGIPPGVINIVSGRGIIVVEATINHPKIKKVAFTGST 247
DB 236 TVVLKTAQTPLSALFASKLHEAGLDPGVVNVVSGFGPTAGAAASHMDVDVKIAFTGST 295
QY 248 ATGRHIYGSAA-AGLKVKVTELEGGKSENIIVFAIAELKKAQONTILGIYVNSGEVCCAGSR 306
DB 296 DTGKVVLEARSNLKSVTLELGGKSPFLIMDDADVDAVELAHFALFFNQGCCAGSR 355
QY 307 VYVEESIYDKFTIEEFKAASESIKVGDFPESTFQAGVTSQMLNKLKYYVDIGKNEGATL 366
DB 356 TVHERIYDEFEKAKALKAPVVGDPFKNTVEGPGIIPGPQNKILPYIKYGVDSGANL 415
QY 367 ITGGERIGSKGYFTKPTVFGVQKFTMPVVKFEIIPGPVTVTKFSADAEVINMANDSEYGL 426
DB 416 VTGGRLGKGYTYQPTTIESVQUNMKIAQEETIIPGPVQSLKFNLDNEVIKRNASQYGL 475
QY 427 AAGIHTSNTALVADRVNAGTVWINTYNDFHVAFPGGFENASGLGREGMSVDALONYLQ 486
DB 476 AAGVTNNINTANTILTRALRVGTVVWNCFFVFAAIFPGDYKSSGSGSEKSIUSLKNYIQ 535

QY 487 VKAV 490
DB 536 VKAV 539

Search completed: June 24, 2003, 10:17:08
Job time: 41 8852 secs

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

697

698

699

700

701

702

703

704

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756

757

758

759

760

761

762

763

764

765

766

767

768

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807

808

809

810

811

812

813

814

815

816

817

818

819

820

821

822

823

824

825

826

827

828

829

830

831

832

833

834

835

836

837

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868

869

870

871

872

873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919

920

921

922

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

967

968

969

970

971

972

973

974

975

976

977

978

979

980

981

982

983

984

985

986

987

988

989

990

991

992

993

994

995

996

997

998

999

1000

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 13:32:35, Search time: 0.2749 seconds

(without alignments)

1696 x3 Million cell updates/sec

Title: US-09-830-751-4

Perfect score: 2625

Sequence: 1 MSAAATCAVAPAPNQQPEVPT...YSGJAYTEVKTIVKVPQKN 500

Scoring table: BLOSUM62

Gap: 10 0, Gapext 0.5

Searched 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A_Geneset_101002.2

```
1: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneset/genesetp-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneset/genesetp-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneset/genesetp-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneset/genesetp-emb1/AA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2625	100.0	500	22	Human aldehyde deh
2	1929	73.5	517	23	Human ALDH5 protei
3	1917	73.0	517	23	Human ALDH5 protei
4	1866	71.1	529	22	Drosophila melanog
5	1771	67.5	412	21	Lung cancer associ
6	1741	66.3	521	15	Aldehyde-dehydroge
7	1718	65.4	512	23	Prostate cancer-as
8	1718	65.4	512	23	Human aldehyde deh
9	1718	65.4	529	22	Novel human diagno
10	1543	58.8	538	21	Arabidopsis aldehy

11	1473	56.1	534	21	AAV67414	Arabidopsis aldehy
12	1470	56.0	534	22	AAB62256	Drosophila melanog
13	1418.5	54.0	497	7	AAP60456	Sequence of aldehy
14	1305.5	49.7	446	16	AAP71841	Clasporium heiba
15	1280	48.8	501	21	AAV62339	Arabidopsis thalia
16	1280	48.8	501	21	AAV67413	Arabidopsis thalia
17	1269	48.3	495	16	AAV71803	Alternaria alterna
18	1225.5	45.7	435	22	AAV74923	Yeast aldehyh-3eh
19	1214	46.2	314	21	AAB56770	Human prostate can
20	1160.5	44.2	523	23	AAB08241	Human aldehyde deh
21	1124.5	42.8	424	21	AAV62240	Arabidopsis thalia
22	1098.5	41.8	909	22	APR58516	Drosophila melanog
23	1022	38.9	357	21	AAV36241	Arabidopsis thalia
24	978.5	37.4	440	22	AAU34454	E. coli cellular p
25	975.5	37.2	496	22	AAU34957	Enterococcus faeci
26	968.5	36.4	440	22	AAU36569	Pseudomonas aerugi
27	962.5	36.7	532	21	AAV35561	Arabidopsis thalia
28	961.5	36.6	503	21	AAV35562	Arabidopsis thalia
29	961.5	36.6	503	21	AAV35517	Arabidopsis thalia
30	957.5	36.5	501	21	AAV16881	Arabidopsis thalia
31	953	36.4	512	22	AAV74926	E. coli aldehyde d
32	950	36.2	497	21	AAC16882	Arabidopsis thalia
33	947.5	36.1	496	22	AAU34148	Staphylococcus aur
34	947.5	36.1	496	22	AAU37064	Staphylococcus aur
35	947	36.1	506	22	AAV45208	V. cholerae VPI pha
36	942.5	35.9	477	22	AAV04880	Micromonospora ova
37	942	35.9	507	22	AAV80035	Pseudomonas putida
38	942	35.9	507	22	AAV80043	Pseudomonas putida
39	939	35.8	446	22	AAV81448	S. epidermidis ope
40	939	35.8	506	23	AAB39348	Staphylococcus epi
41	939	35.8	507	23	AAB48400	Zoysia tenuifolia
42	933.5	35.6	517	23	AAB41923	Human ovarian anti
43	930	35.4	508	21	AAV02082	Aldehyde dehydroge
44	930	35.4	508	23	AAV16744	Phodococcus erythr
45	925	35.3	485	23	AAV15196	Phodococcus picric

ALIGNMENTS

RESULT 1

AAV74924

ID AAB74924 standard; Protein: 500 AA.

AC AAB74924;

XX

XX

DE 25-JUN-2001 (first entry)

DE Human aldehyde dehydrogenase ALDH2 protein sequence SEQ ID NO:4.

XX Aldehyde dehydrogenase, glycerol dehydratase, 3-HP, glycerol;

XX feedstock; 3-hydroxypropionic acid; genetic engineering; glucose;

XX bacterial host; absorbable prosthetic device; surgical suture;

XX beta-lactam; acrylic acid; trifluoromethylated alcohol; diol;

XX polyhydroxyalkonate; copolymer; lactic acid.

XX Homo sapiens.

XX

XX WO200116346-A1.

XX

XX 08-MAR-2001.

XX

XX 30-AUG-2000; 2000U-US23878.

XX

XX 30-AUG-1999, 99US-0151440.

XX (WISC) WISCONSIN ALUMNI REFS FOUND.

XX Suthers PF, Cameron DC;

XX WPI, 2001-315988/33.

XX N-PSDB; AAF20082.

XX

PT 3-Hydroxypropionic acid preparation, for use e.g. as monomer, by
 PT fermenting recombinant microorganisms expressing genes for suitable
 PT enzymes in the presence of glycerol or glucose -

XX Claim 5; Page 32-35; 63pp; English.

XX The present invention describes a method for the production of
 CC 3-hydroxypropionic acid (3-HP). The method comprises fermenting a
 CC recombinant microorganism in the presence of a source of glycerol, (i)
 CC or glucose, where the microorganism: (i) expresses genes for non-native
 CC enzymes which catalyse the production of (3-HP) from (i); (ii) carries
 CC genetic constructions for the expression of a glycerol dehydratase
 CC (GDH) and aldehyde dehydrogenase (ADH) capable of catalysing the
 CC production of (3-HP) from (i); or (iii) carries a genetic construct
 CC which expresses the dhap gene from Klebsiella pneumoniae and a gene for
 CC an ADH capable of catalysing the production of (3-HP) from (i). 3-HP is
 CC a monomer, and is useful e.g. in the production of absorbable prosthetic
 CC devices and surgical sutures or for incorporation into beta-lactams,
 CC production of acrylic acid or formation of trifluoromethylated alcohols
 CC or diols, polyhydroxyalkonates and copolymers with lactic acid.
 CC Incorporation of genes encoding two enzymes makes the host organisms
 CC able to produce (3-HP) from (i). The biotechnological method of
 CC preparing (3-HP) is potentially cheaper than chemical synthesis. The
 CC present sequence represents the human aldehyde dehydrogenase ALDH2, which
 CC is used in the exemplification of the present invention.

XX Sequence 500 AA;

Query Match 100.0%; Score 2625; DB 22; Length 500;
 Best Local Similarity 100.0%; Pred. No. 6e-239;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSAATQAVPAPQOPEVFCNOIFINNEWHDAVSRRKIFPTVNPSTGEVLCVVAEGSKEDV 60
 DB 1 MSAATQAVPAPQOPEVFCNQIFINNEWHDAVSRRKIFPTVNPSTGEVLCVVAEGSKEDV 60
 QY 61 DKAREGPAGFQGLSPWRMDASHSGRLNRLADLIERDRTYLAALFETLDNGKPVVSYL 120
 DB 61 DKAREGPAGFQGLSPWRMDASHSGRLNRLADLIERDRTYLAALFETLDNGKPVVSYL 120
 QY 121 VLDLMLKCLRYVAGWADKYHCKTIPIDGDFSYTPHPVGVCTGCIIPWNPFLLMQAWKL 180
 DB 121 VLDLMLKCLRYVAGWADKYHCKTIPIDGDFSYTHIEPVGVCGCIIPWNPFLLMQAWKL 180
 QY 181 GPALATGNVVMKVAEQTPITALYVANLLKEAGFPVGVNIVPFGPTAGAAIASHESVD 240
 DB 181 GPALATGNVVMKVAEQTPITALYVANLLKEAGFPVGVNIVPFGPTAGAAIASHESVD 240
 QY 241 KVAFTGSTRIGRVIOVAGSSNLRKVTLELGGKSPNIDMSQADMWAVEQAHFALFNQGS 300
 DB 241 KVAFTGSTRIGRVIOVAGSSNLRKVTLELGGKSPNIDMSQADMWAVEQAHFALFNQGS 300
 QY 301 QCCAGSKRTVQEDLYDEFWASVAKAKSVGNPFUSKIFQSGPVVDETQFKKILGYINT 360
 DB 301 QCCAGSKRTVQEDLYDEFWASVAKAKSVGNPFUSKIFQSGPVVDETQFKKILGYINT 360
 QY 361 GKQEGAKLGGGIAADRGYFIQPTVGVIVQDGMTIAKEETPGVMQILKEKTEIEVVGR 420
 DB 361 GKQEGAKLGGGIAADRGYFIQPTVGVIVQDGMTIAKEETPGVMQILKEKTEIEVVGR 420
 QY 421 ANNSYGLAAAVFTKILUKANYLSQAQAGTWVWNYGVFGAGSPGCKYKSGSGPTPLGE 480
 DB 421 ANNSTYGLAAAVFTKILUKANYLSQAQAGTWVWNYGVFGAGSPGCKYKSGSGPTPLGE 480
 QY 481 YGLQAYTEVKTVTKVPQKN 500
 DB 481 YGLQAYTEVKTVTKVPQKN 500

RESULT 2

ID AAM49516 standard; Protein; 517 AA.

XX

AC

AAM49516;

XX 13-MAY-2002 (first entry)

XX Human ALDH5 protein.

XX

KW ALDH5; human; polymorphism; haplotype; aldehyde dehydrogenase 5;
 KW binding affinity; drug targeting; alcoholism; alcohol induced disorder;
 KW antialcoholic.

XX Homo sapiens.

XX W0200192279-A2.

XX 06-DEC-2001.

XX 29-MAY-2001; 2901W-0817253.

XX 26-MAY-2000; 240038-20750HP.

XX (GENA-) GENA1SSANTIE FIARM (N°).

XX Duda A, Finkel K, Kazemi A, Messer C, Sanchez A;

XX W01; 2002-122054/16.

XX N-PSDH; ARA99282, ARA99283.

XX

PT New genetic variants with polymorphisms in the aldehyde dehydrogenase 5
 PT (ALDH5) gene, useful for studying the function of ALDH5, and for
 PT expressing ALDH5 protein which is useful in screening drugs for
 PT treating ALDH5 related diseases -
 XX Claim 40; page 72-74; 96pp; English.

PS

CC This invention describes a novel isolated genes and haplotypes of the
 CC human aldehyde dehydrogenase 5 (ALDH5) gene containing polymorphic sites.
 CC The polymorphic ALDH5 variant is useful in studying the effect of the
 CC variation on the biological activity of ALDH5 and on the binding
 CC affinity of candidate drugs targeting ALDH5 for the treatment of
 CC alcoholism and alcohol-induced disorders. Polynucleotides comprising a
 CC polymorphic gene variant or fragment may be used for therapeutic
 CC purposes. ALDH5 protein isoforms may be used in assays to measure the
 CC binding affinities of one or more candidate drugs targeting the ALDH5
 CC protein. ALDH5 proteins may be used to generate antibodies. Haplotyping
 CC method can be used by scientists to validate ALDH5 as a candidate
 CC target for treating a specific condition or disease predicted to be
 CC associated with ALDH5 activity, and in the design of clinical trials of
 CC candidate drugs for treating a specific condition or disease predicted
 CC to be associated with ALDH5 activity. Information on polymorphisms on
 CC the ALDH5 gene can be applied for studying the biological function of
 CC ALDH5 as well as in identifying drugs targeting this protein for the
 CC treatment of disorders related to its abnormal expression or function.
 CC The products of the invention have antialcoholic activity. This sequence
 CC represents the human ALDH5 protein described in the disclosure of the
 CC invention.

XX Sequence 517 AA;

SQ

Query Match 73.5%; Score 1929; DB 24; Length 517;
 Best Local Similarity 72.7%; Pred. No. 4.5e-174;
 Matches 363; Conservative 61; Mismatches 74; Indels 2; Gaps 1;

QY 2 SAATQAVPAPQOPEVFCNPQINNEWHDAVSRRKIFPTVNPSTGEVLCVVAEGSKEDV 61

DB 20 SSAA--ALPSPITNPPIYNQJFINNEWDVASEKRTVNTTTEVTGHWAEGRALV 77

QY 62 KAREGRCAGFQGLSPWRMDASHSGRLNRLADLIERDRTYLAALFETLDNGKPVVSYL 121

DB 78 PAVKAAPEAFRIGSPWPPMDASERPTLADVFGDPVYLAFTLIDNGKPEQSTAL 137

QY 122 DLIMVLKCLPYVAGWADKYHCKTIPIDGDFSYTHIEPVGVCTGCIIPWNPFLLMQAWKL 181

DB 138 DLIDEVTKVRYVAGWADKYHCKTIDMHGQHCFTTTHIEPVGVCTGCIIPWNPFLLMQAWKL 197

AC ABB60140.
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polyprotein; cell signalling; insecticide;
 XX
 KW Drosophila, developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN W0200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-0509231.
 XX
 PR 23-MAR-2000; 2000US 1916371.
 XX
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 XX
 DR N-PSDB; ABL04243.
 XX
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 7212; 21pp + Sequence Listing, English.
 XX
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 520 AA:
 Query Match 71.1%; Score 1866; DB 22; Length 520;
 Best Local Similarity 69.8%; Pred. No. 3,1e-167;
 Matches 351; Conservative 61; Mismatches 8; Indels 8; Gaps 3;
 QY 2 SAAATQAVPAPNQOPPEVFCNQIFINNEHIDAVSRKTEPTVNPSTGEVCGVAEGDKEDVD 61
 DB 1 AVANYSLLPQPTTIDLYTGVFINNEHKSCKIFETINPTAEVIAETIQACKEDID 80
 QY 62 KAPPEHPVAFOLCSWPRMWDASHSCILNPLADLIERDFTYLAETLDNCKFYVISLY 121
 DB 81 IAVQAAARNAPKLGSPWRMRDASERGLLYRIADLIMERDQVYLASLETIDNCKPYSMSYNV 140
 QY 122 DLDMLVKCLRYAGWADKYHGKTIPIIDGFTSYTRHEPVGCGQIFPWNFPLLMCAWKLK 181
 DB 141 DLPFAIKNLYFAGWADKNHCKTIPMDCDFTYTRHEPVGCGQIIFWNFPILLMAWKLK 200
 QY 182 PALATGNVVMVKAQETPLTALYVANLKEAGFPFGVNVNIPVPGFTAGTAIAHSHEDVDK 241
 DB 201 PALATGNTIVLKPAQETSLTALYTAQIVKFAEPFGVNVNIPVPGFTAGTAIAHSHEDVDK 254
 QY 242 VAFTGSTEIGRVIOVAAGSSNLKRYTLELGGKSPNIMSDALMDWAVEQAHFALFFNQCG 301
 DB 260 VAFTGSTEIGRVIOVAAGSSNLKRYTLELGGKSPNIMSDALMDWAVEQAHFALFFNQCG 319
 QY 302 CCGAGSPTIVQRTIVDFEVRVSVAPAKSRVGNPFDSEKTEGPOVDTEPKKIIIGYINTG 361
 DB 361 CCGAGSPTIVQRTIVDFEVRVSVAPAKSRVGNPFDSEKTEGPOVDTEPKKIIIGYINTG 361

320 CCGAGSPTIVQRTIVDFEVRVSVAPAKSRVGNPFDSEKTEGPOVDTEPKKIIIGYINTG 379
 362 KQEGAKLVAG3SKPEGLP---GYFVQPIVFAVQVQMIIAKEETIPGVQVQIIKEETIDEV 419
 389 KQEGAKLVAG3SKPEGLP---GYFVQPIVFAVQVQMIIAKEETIPGVQVQIIKEETIDEV 436
 418 VUKANNSTVGLAAAVFELKULOKANYLSUALQAIIWVWVYDVFCAQSTGEGYKNSQSKRE 477
 437 IERANNSEYGLAAAVFETKIDUKANYIVGELKAGTIVWNTYNNVLAQAIFEGYKNSQSKRE 496
 478 IAFYGLQAYTEVKTIVTKVPOKN 500
 497 NGFYALSNYTEVKSIVKVAUKN 519

RESULT 5
 AAB58398
 ID AAB58398 standard; Protein; 412 AA.
 XX
 AC AAB58398;
 XX
 DT 14-MAR-2001 (first entry)
 XX
 DE lung cancer associated polypeptide sequence SEQ ID 736.
 XX
 KW Human, lung cancer associated protein; neuroprotective; cytostatic;
 KW cardioactive; immunomodulatory; muscular active; vulnerary;
 KW gastrointestinal; nephrotropic; antineoplastic; syncretological;
 KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
 KW proliferative disorder; wound healing; infectious disease.
 XX
 OS Homo sapiens.
 XX
 PN W0200055180-A2.
 XX
 PD 21-SEP-2000.
 XX
 PR 08-MAR-2000; 2000WO-0505918.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (KOSK/) KOSEN C A.
 XX
 PI Ruben SM;
 XX
 WIPI: 2000-587514/55.
 N-PSDB: AAB18274.
 PT lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer.
 XX
 PS Claim 11; Page 1253-1254; 1425pp; English.
 XX
 CC Polynucleotide sequences AAF17922 AAF18423 encode human lung cancer
 CC associated proteins represented in AAB58106 AAB58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective, cytostatic, cardioactive,
 CC immunomodulatory, muscular active general, vulnerary, gastrointestinal
 CC general, nephrotropic, antineoplastic, syncretological, or antibacterial
 CC activity. The invention also includes antibodies specific for the
 CC protein or polynucleotide sequences. The lung cancer associated
 CC polynucleotide sequences may be used for detection of lung cancer,
 CC chromosome identification, as chromosome markers, and for numerous other
 CC diagnostic or research purposes. The proteins may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders. The proteins may also be used in the treatment of wounds and
 CC infectious diseases. Polynucleotide sequences AAF18425 AAF18434 and
 CC peptide AAB58549 are used in the course of the invention for the
 CC identification and characterisation of the polynucleotide and protein
 CC sequences.


```

XX      SQ      Sequence      412 AA;
XX
XX      Query Match
XX      Best Local Similarity      67.59, Score 1777, DR 21, Length 412,
XX      Matches 339, Conservative 7, Mismatches 28, Indels 0, Gaps 0,
XX
XX      QY      1 MSAATQAVAPAPNOOPEVFNQIFINNEHDAVSKKIFETVNPSTGEVLCQVABGKELV 60
XX      DB      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX
XX      QY      61 DKAREGRGAFQIGSPWRMDASHGRLNKLADLIERDRTYLAALLETIDNGKPYVISYL 120
XX      DB      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX
XX      QY      80 DKAVKAAPAAQIGSPWPPMDASHGRLNKLADLIERDRTYLAALLETIDNGKPYVISYL 139
XX      DB      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX
XX      QY      121 VOLDMLVKLRYAGADKADYHGKTIPIDGDFSYTHREPVGVGCGIIPWNPFLMQAWKL 180
XX      DB      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX
XX      QY      140 VOLDMLVKLRYAGADKADYHGKTIPIDGDFSYTHREPVGVGCGIIPWNPFLMQAWKL 199
XX      DB      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX
XX      QY      181 GPALATGNVVMKVAEOTPTLTALYVANLIKEAGFPVGVNIVPGPPTAGAAIASHEDVD 240
XX      DB      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX
XX      QY      200 GPALATGNVVMKVAEOTPTLTALYVANLIKEAGFPVGVNIVPGPPTAGAAIASHEDVD 259
XX      DB      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX
XX      QY      241 KVAFGTSTETGVLCVAGSSNLKRVTLLELGSKSPNIIIMSDALIMWAVHQAHLAFNAG 300
XX      DB      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX
XX      QY      260 KVAFGTSTETGVLCVAGSSNLKRVTLLELGSKSPNIIIMSDALIMWAVHQAHLAFNAG 319
XX      DB      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX
XX      QY      301 QCCAGSGTEFVQEDYDFVFPVAPAKSHVNVNPFESKTELELQVLETPFKKILISYINT 360
XX      DB      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX
XX      QY      320 QCCAGSGTEFVQEDYDFVFPVAPAKSHVNVNPFESKTELELQVLETPFKKILISYINT 379
XX      DB      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX
XX      QY      361 QKQEGAKILCGGI 374
XX      DB      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
XX
XX      QY      380 GSKRGRCVVVGAI 393
XX      DB      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
XX
XX      RESULT 6
XX      ID      AAP63672 standard; Protein; 521 AA
XX      AC      AAP63672;
XX      DT      08-MAY-1995 (first entry)
XX      DE      Aldehyde-dehydrogenase Aldh-1.
XX      KW      Retro virus; vector; aldehyde-dehydrogenase;
XX      KW      glutamylcysteine-synthetase; hematopoietic cell; cyclophosphamide;
XX      KW      chemotherapy; transgenic animal; gene therapy; cancer therapy;
XX      KW      selectable marker.
XX      OS      Homo sapiens.
XX      PN      W09423015-A.
XX      PD      13-OCT-1994.
XX      PF      01-APP-1994; 94WO-050624
XX      PR      01-APP-1993; 94US-0941722.
XX      PA      (UYCO ) UNIV COLUMBIA NEW YORK.
XX      PI      Dalla-favera R, Gianni AM;
XX      DR      WPI; 1994-333177/41.
XX      DR      N-PSDB; AAQ72450.
XX      PT      Retroviral vectors encoding human cytosolic aldehyde
XX      PT      dehydrogenase or glutamyl cysteine synthetase - used to transform
XX      PT      a subject's haematopoietic cells to reduce the toxic effects of
XX      PT      cyclo phosphamide chemotherapy
XX

```

```

PS      Disclosure; Fig.4; 92pp; English.
XX
XX      CC      A novel retro virus vector encodes human cytosolic aldehyde-
XX      dehydrogenase and/or human glutamylcysteine-synthetase (AAP63673).
XX      Hematopoietic cells transfected by the vector are resistant to
XX      cyclophosphamide, providing a means of gene therapy that allows
XX      higher doses of toxic drugs to be used in cancer chemotherapy.
XX      The human genes may also be used as selectable markers for
XX      mammalian cell transfection and for transgenic animal breeding.
XX
XX      SQ      Sequence      521 AA;
XX
XX      Query Match
XX      Best Local Similarity      66.38, Score 1741, DR 15, Length 521,
XX      Matches 332, Conservative 69, Mismatches 99, Indels 4, Gaps 2;
XX
XX      QY      1 MSAATQAVAPAPNOOPEVFNQIFINNEHDAVSKKIFETVNPSTGEVLCQVABGKELV 60
XX      DB      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX
XX      QY      61 DKAREGRGAFQIGSPWRMDASHGRLNKLADLIERDRTYLAALLETIDNGKPYVISYL 118
XX      DB      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX
XX      QY      77 DKAVKAAPAAQIGSPWPPMDASHGRLNKLADLIERDRTYLAALLETIDNGKPYVISYL 136
XX      DB      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX
XX      QY      119 YLVLLIMVLKELPYAGAWALKY--HKKILIPLEGEFSYTHREPVGVGCGIIPWNPFLMQ 176
XX      DB      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX
XX      QY      137 YLVLLIMVLKELPYAGAWALKY--HKKILIPLEGEFSYTHREPVGVGCGIIPWNPFLMQ 196
XX      DB      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX
XX      QY      177 AWKIGPALATGNVVMKVAEOTPTLTALYVANLIKEAGFPVGVNIVPGPPTAGAAIASH 246
XX      DB      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX
XX      QY      197 LWKIGPALATGNVVMKVAEOTPTLTALYVANLIKEAGFPVGVNIVPGPPTAGAAIASH 256
XX      DB      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX
XX      QY      237 ELVLLIMVLKELPYAGAWALKY--HKKILIPLEGEFSYTHREPVGVGCGIIPWNPFLMQ 296
XX      DB      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX
XX      QY      257 WMDKVAFTGSTGVKLIKEAGSNLKRVTLELGSKSPNIIIMSDALIMWAVHQAHLAFNAG 316
XX      DB      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX
XX      QY      297 FNOQCCAGSGTEFVQEDYDFVFPVAPAKSHVNVNPFESKTELELQVLETPFKKILISYINT 356
XX      DB      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX
XX      QY      317 YHQQGCGTAAASPIPVRESLYDEEFPVPSVERAKRYILGNPLTPGATGSPQIDKQYGRKILID 376
XX      DB      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX
XX      QY      357 YINTGKQEGAKILCGGI 374
XX      DB      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
XX
XX      QY      377 LIESGKRGAKLECGGPGWENKGVFQPTVFSNVTDEMPIAKEFIIEPVDQIMKFKSIDP 436
XX      DB      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX
XX      QY      417 VVGRANSTYGLAAAVFTKDLKANKYLSQALQAGTVWVNCYDFGAGSPFGYKMSGGR 476
XX      DB      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX
XX      QY      437 VIKFANNIFGSLASGVFLKDLKALITSSALAGTVMVNYGVVSACTPFGCFKMSGNCP 496
XX      DB      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX
XX      QY      477 ELGEYGLQATEVKTIVTKVPQKN 500
XX      DB      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX
XX      QY      497 ELGEYGFREYTKIVTKISQKN 520
XX      DB      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX
XX      RESULT 7
XX      ID      ABC61842 standard; Protein; 512 AA.
XX      AC      ABC61842;
XX      DT      15-AUG-2002 (first entry)
XX      DE      Prostate cancer-associated protein #43.
XX      KW      Prostate cancer; prostate tumour tissue; human; mammal; cytostatic.
XX      OS      Mammalia.
XX      PN      W0200230268-A2.
XX      PF      18-APP-2002.
XX      PR      12-OCT-2001; 2001WO-0542045.
XX

```

13-OCT-2000; 2000US-0687576.
 08-DEC-2000; 2000US-0733288.
 08-DEC-2000; 2000US-0733742.
 24-JAN-2001; 2001US-2539742.
 16-MAR-2001; 2001US-276791P.
 16-MAR-2001; 2001US-276888P.
 06-APR-2001; 2001US-281922P.
 24-APR-2001; 2001US-286214P.
 30-APR-2001; 2001US-0847046.
 04-MAY-2001; 2001US-288584P.
 (EOSB-) FOS BIOTECHNOLOGY INC.
 Gish KC, Mack DH, Wilson KE, Afar D, Hevezi P;
 WPI; 2002-471345/50.
 N-PSDB: ABK92157.
 Detecting a prostate cancer-associated transcript in a cell in a patient, useful for diagnosing prostate cancer (PC) or screening modulators of PC, by determining if prostate cancer-associated genes are expressed in a prostate tissue.
 Claim 27; Page 333-334; 436pp, English.
 The present invention relates to methods of detecting a prostate cancer-associated transcript in a cell from a patient. The method comprises contacting a biological sample from the patient with prostate cancer-associated polynucleotides (designated PC genes) that selectively hybridize to a sequence that is at least 80% identical to them. The prostate cancer-associated polynucleotide sequences are differentially expressed in prostate tumour tissue or in prostate cancer and are derived from the tissues of various organisms such as humans or other mammals (e.g. mice, sheep and dogs). The methods of the invention are useful for diagnosing and treating prostate cancer in mammals. The prostate cancer-associated genes are useful for diagnosing or treating prostate cancer, as well as for identifying modulators of prostate cancer or agents that inhibit prostate cancer. The nucleic acid sequences are particularly useful in gene therapy, as a vaccine or in antisense applications.
 AB61800-AB61944 represent prostate cancer-associated proteins.

XX Sequence 512 AA;

Query Match 65.4%; Score 1718; PR 23; Length 512;
 Best Local Similarity 65.5%; Pred. No. 2.9e-153;
 Matches 324; Conservative 66; Mismatches 104; Indels 0; Gaps 0;
 8 AVPAIPNOQPEVFCNOIFINNEWHDAVSRKTFPTVNPSTGCVICQVAREGDKEDVKAHGR 67
 19 ALPRIRNLEVRKTRIFINNEWHESKSKKFKATCNPSTREQICEVEFGDKEDVKAVERA 78
 68 PTAFQIGSPWPMWDASHSGRIINPLADIPRPTVIAAEITLNGKTPVVISLVLELWVI 127
 79 QVAFORGSPWRLLDALSGRLQLADLVERDRATLAELTMDTKPPLHAFIDLECCI 138
 128 KCLRYTAGNADYHGKTIPIGCGFTSYTHHEPVGNGGIIIPWNIPLMAWKLSPALIG 187
 139 RILRYFAGWADKIGKTIPTTGNVVFETPHEPFGVSGAIIIPWNEPFLMLVWKLAPLCCG 198
 188 NVVMKVAEOTPTLTALYVANLKEAGPPGVVNIIVPGPTAGAAIASHEVDVAVATGS 247
 199 NTWLPAPAEPTLTALYVSLIKKAGPPGVVNIIVPGPTVGAANISHPDINPIATTS 254
 248 TELGRVIOVAASNNIKPVTLFLGKSPNIIMSPADMWAVEQAEALFFNQGQCCAGS 307
 259 TEVGKIVAKKAASRNLKRVTLGLGKNPICADADLDAVECAHOGVFNQGCCTAAS 318
 308 RTEVQEDIDYDFWRSVAKSRVYGNPDSKTEGPOVDTPFKKILGYINTQKQCAK 367
 319 RVFEQGVYSEVRVYAKRPVGDPTVKTEGQGPQKQKFKITLFTESKKECAK 378
 368 LLCGGIAADRGYFIQPTVFGDVGDMTIAKEEIFGPVWVILKFTITEEVVGRANNSTYG 427

DB 379 LECGSGAMEDKGLFKPTVFSEVTDNMRIAKETIEGTVQPIILKPKSIEVTKRANSHYG 438
 QY 428 LAAAVETKDLDRKANLSCALJASIVWVN-YLVIKAGSGEGYKMSGSGREJEGELCAYT 487
 DB 439 LTAAVFTKNLDKALKLASALESGTVINCYNALYQAAPFGGPKMSGNGKELGAYALAYT 498
 QY 488 EYKTVTVKVPQKN 500
 DB 499 EYKTVTIELEDPN 511

RESULT 8

AA017364
 ID AA017364 standard; protein: 512 AA.

XX AC AA017364;

XX 19-JUL-2002 (first entry)

XX Human aldehyde dehydrogenase 6.

XX Human; endometriosis; DNA chip; fibronectin; p27; reticulocalbin;
 KW aldehyde dehydrogenase 6; gravin; phospholipase C epsilon; elastin;
 KW insulin-like growth factor binding protein-2; alpha-2 type IV collagen;
 KW transmembrane receptor pTK7; collagen type XVIII alpha 1;
 KW platelet derived growth factor receptor alpha; laminin M chain;
 KW subtilisin like protein PACE4; nidogen.

XX Homo sapiens.

XX EP1191107-A2.

XX 27-MAP-2002.

XX 21-AUG-2001; 2001EP-0250300.

XX 25-SEP-2000; 2000DE-1048633.

XX (SCHD) SCHERING AG.

XX Hess-stumpff H, Haendler H, Kraetzschmar J, Kreft B, Winterhaup B;
 Regidor F, Scotti S;

XX WPI; 2002-317413/36.

XX In vitro diagnosis and monitoring of endometriosis, comprises

XX detecting reduced expression of specific gene products, e.g. from the

XX fibronectin gene.

XX Claim 1; Page 17-18; 21pp; German.

XX The present invention relates to a method for the in vitro diagnosis of
 XX endometriosis by determining the amount of gene product from at least one
 XX specific gene in a patient sample and comparing this with the amount of
 XX gene product in a control sample. A reduced level is indicative of
 XX endometriosis. The gene products may be fibronectin, p27, reticulocalbin,
 XX aldehyde dehydrogenase 6, gravin, phospholipase C epsilon, elastin,
 XX insulin-like growth factor binding protein-2, alpha-2 type IV collagen,
 XX transmembrane receptor pTK7, collagen type XVIII alpha 1,
 XX platelet derived growth factor receptor alpha, laminin M chain.

XX subtilisin like protein PACE4 or nidogen. The method is useful for
 XX initial diagnosis of endometriosis, and also for monitoring progress and
 XX treatment of the disease. The present sequence is human aldehyde
 XX dehydrogenase 6.

XX Sequence 512 AA;

Query Match 65.4%; Score 1718; PR 23; Length 512;

Best Local Similarity 65.5%; Pred. No. 2.9e-153;

Matches 324; Conservative 66; Mismatches 104; Indels 0; Gaps 0;

QY 8 AVPAIPNOQPEVFCNOIFINNEWHDAVSRKTFPTVNPSTGCVICQVAREGDKEDVKAHGR 67

```

Db      19 ALPRIRNLVKKFTKFINNEHESKGGKFAFCNSTRPQICEVEEGKPPVUKAVEAA 78
QY      68 PGAFQLSGPWRMDASGRLLNPIADLTERPTYIAAETLNGSKPYVISLVLDLMDVL 127
Db      79 QVAFQSGPWRRLDALSRGLLHQLADLVERDRATLAALETMDTGKPPFLHAFIDLEGCI 138
QY      128 KCLRYAGWADKYHGKTIPIDGDFFSYTRHEPVGSGSIIPNNFPLLMQAWKLGALATG 187
Db      139 RILRYFAGWADKIQKTIPIIDUNVCFTRHEPVGSGSIIPNNFPLLMQAWKLGALATG 198
QY      188 NVVMKVAEQPTLITALLYANLKEAGPPGVVNIIVFGFGPTAGAAIASHEDYKVAFTGS 247
Db      199 NTMLKDAEQPTLITALLYLGLSLKEAGPPGVVNIIVFGFGPTVGAALSSHPQINKIAFTGS 258
QY      248 TEIGRVITQVAGSSNIKPVTIETGGSFNLIIMSDALMDWAVEGAHFALEFNGQCCAGS 307
Db      259 TEVSKLVKEAASRNKRPVTEIGGKNPILVWALADLDAVFAHGVFFNGQCCAGS 318
QY      308 FTFVQETIYDFEVVPSVAPAKSPVVGNFPSFSTEQPQVETGFKKILSYINFRKQSAK 367
Db      319 RVFVEEQVYSFVRSVEYAKKRPVGLPQVVKTEGHPQIQKQFUKLLEIESGKKEGAK 378
QY      368 LAFSSSTAAGPSEYELQTVRQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 427
Db      379 LECGGSAMENKGLFKPTVFSEVTDNNPTAKRFTPGVPQVILKFKSIEEVIPKASTDYG 438
QY      428 LAAAVFTKLDKALKLASALESGTIVWVNCYVDFVTAQSPDDYKMSISGPELGYGLAYT 487
Db      439 LTAAVFTKLDKALKLASALESGTIVWVNCYVDFVTAQSPDDYKMSISGPELGYGLAYT 498
QY      488 EVKTVTVKVPQKN 500
Db      499 EVKTVTVKLGDKN 511

```

RESULT 9

ABG06577
ID ABG06577 standard; protein: 529 AA.

AC ABG06577;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #6568.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
food supplement, medical imaging, diagnostic, genetic disorder

OS Homo sapiens.

PN WO2001/5067 A2.

PO 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0649167.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS0764.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

PS claim 20; SEQ ID NO 364a; lang: English

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligonucleotides, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (I) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. Abgenovo-Archiv77 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pat_sequences.

XX Sequence 529 AA:

```

Query Match      65.4%  Score 1718;  DB 22;  Length 529;
Best Local Similarity 65.5%  Pred. No. 31e-153;
Matches 424;  Conservative 66;  Mismatches 104;  Indels 0;  Gaps 0;
QY  H AVPAWNPQPEVFNQIPIFNNHNIIVSPKTIPIVNFSTIGSEVILVAEAGKHIVLRAFPSP 67
Db  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY  36 ALPRIRNLVKKFTKIFINNEHESKGGKFAFCNSTRPQICEVEEGKPPVUKAVEAA 95
Db  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY  68 PGAFQLSGPWRMDASGRLLNPIADLTERPTYIAAETLNGSKPYVISLVLDLMDVL 127
Db  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY  96 QVAFQSGPWRRLDALSRGLLHQLADLVERDRATLAALETMDTGKPPFLHAFIDLEGCI 155
Db  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY  128 KCLRYAGWADKYHGKTIPIDGDFFSYTRHEPVGSGSIIPNNFPLLMQAWKLGALATG 187
Db  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY  156 RILRYFAGWADKIQKTIPIIDUNVCFTRHEPVGSGSIIPNNFPLLMQAWKLGALATG 215
Db  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY  188 NVVMKVAEQPTLITALLYANLKEAGPPGVVNIIVFGFGPTAGAAIASHEDYKVAFTGS 247
Db  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY  216 TEIGRVITQVAGSSNIKPVTIETGGSFNLIIMSDALMDWAVEGAHFALEFNGQCCAGS 275
Db  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY  248 TEIGRVITQVAGSSNIKPVTIETGGSFNLIIMSDALMDWAVEGAHFALEFNGQCCAGS 307
Db  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY  276 TEVSKLVKEAASRNKRPVTEIGGKNPILVWALADLDAVFAHGVFFNGQCCAGS 335
Db  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY  308 FTFVQETIYDFEVVPSVAPAKSPVVGNFPSFSTEQPQVETGFKKILSYINFRKQSAK 367
Db  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY  336 RVFVEEQVYSFVRSVEYAKKRPVGLPQVVKTEGHPQIQKQFUKLLEIESGKKEGAK 395
Db  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY  368 LAFSSSTAAGPSEYELQTVRQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 427
Db  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY  396 LECGGSAMENKGLFKPTVFSEVTDNNPTAKRFTPGVPQVILKFKSIEEVIPKASTDYG 455
Db  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY  428 LAAAVFTKLDKALKLASALESGTIVWVNCYVDFVTAQSPDDYKMSISGPELGYGLAYT 487
Db  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY  456 LTAAVFTKLDKALKLASALESGTIVWVNCYVDFVTAQSPDDYKMSISGPELGYGLAYT 515
Db  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY  488 EVKTVTVKVPQKN 500
Db  1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY  516 EVKTVTVKLGDKN 528

```

RESULT 10

AA67412
ID AA67412 standard; protein: 538 AA.

XX AA67412;

XX 12-MAY-2000 (first entry)


```

Db      197  LVASLVREAGPPGVINVISOPFVACAAALSSHMPTVKVAFGTSTVVGFTILKAASSN 256
QY      263  LKRVILEGGKSPINILMSIALMDWAVEAHFALFFNQCCQCCAGSRFFVUEDIYDFVVR 322
Db      257  LKKVILEGGKSPNIVFEDADIDNAISWVNFGEFNNHQQCCQCCAGSRVYVJESLYDFVQK 316
QY      323  SVARAKSRVVGSPFUSKTEQSPQDETQFKKILGY INQKQESAKLLGQSTIAADPGYFI 382
Db      317  PFERAKNVVGVIPFAALIFQAPQSKVGFQFIMFY LQARKLQATVETSPKSGKSGYFI 476
QY      383  QPTVEGDVDGMTIAKRETFPGVMQILKFKTIEEVGFRANNSTYGLAAAVFTKDLQKANY 442
Db      377  EPTIE-SNVTEDMKIVKEEIPGPVCSIAKFKTKEDAIKLGNASIVYGLAAAVHTKLNLTALE 436
QY      443  LSAALQALGVWVNYVDFVCAVSPPGSKYMSISGFFLQYGLLQAYTEVKTVTVAV 496
Db      437  VSNLAKAGTVWVNTYTLHHQMPGGYKESGIGRELAEDALANTYTKTKVTSIRL 490

RESULT 15
AAG36239
ID   AAG36239 standard, Protein, 501 AA.
XX
AC   AAG36239;
CV   18-OCT-2000 (first entry)
XX
DF   Arabidopsis thaliana protein fragment SEQ ID NO: 44363.
XX
KW   Protein id-identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000
XX
PF   25-FEB-2000; 2000EF-0301430
XX
PF   25-FEB-1999; 99US-0121825
PF   05-MAP-1999; 99US-0123180
PF   09-MAP-1999; 99US-0123548
PF   23-MAR-1999; 99US-0125788
PF   29-MAP-1999; 99US-0126264
PF   29-MAP-1999; 99US-0126785
PF   01-APR-1999; 99US-0127462
PF   06-APR-1999; 99US-0128234
PF   08-APR-1999; 99US-0128714
PF   16-APR-1999; 99US-0129845
PF   19-APR-1999; 99US-0130077
PF   21-APR-1999; 99US-0130444
PF   23-APR-1999; 99US-0130510
PF   23-APR-1999; 99US-0130801
PF   30-APR-1999; 99US-0131449
PF   30-APR-1999; 99US-0132048
PF   04-MAY-1999; 99US-0132407
PF   04-MAY-1999; 99US-0132484
PF   05-MAY-1999; 99US-0132485
PF   06-MAY-1999; 99US-0132486
PF   06-MAY-1999; 99US-0132487
PF   07-MAY-1999; 99US-0132863
PF   11-MAY-1999; 99US-0134256
PF   14-MAY-1999; 99US-0134218
PF   14-MAY-1999; 99US-0134219
PF   14-MAY-1999; 99US-0134221
PF   14-MAY-1999; 99US-0134370
PF   18-MAY-1999; 99US-0134768
PF   19-MAY-1999; 99US-0134941
PF   20-MAY-1999; 99US-0135124
PF   21-MAY-1999; 99US-0135353
PR      24-MAY-1999; 99US-0135624
PR      26-MAY-1999; 99US-0136021
PR      27-MAY-1999; 99US-0136392
PR      28-MAY-1999; 99US-0136782
PR      01-JUN-1999; 99US-0137222
PR      03-JUN-1999; 99US-0137528
PR      04-JUN-1999; 99US-0137502
PR      07-JUN-1999; 99US-0137724
PR      08-JUN-1999; 99US-0138094
PR      10-JUN-1999; 99US-0138540
PR      10-JUN-1999; 99US-0138847
PR      14-JUN-1999; 99US-0139119
PR      16-JUN-1999; 99US-0139452
PR      16-JUN-1999; 99US-0139453
PR      17-JUN-1999; 99US-0139492
PR      18-JUN-1999; 99US-0139454
PR      18-JUN-1999; 99US-0139455
PR      18-JUN-1999; 99US-0139456
PR      18-JUN-1999; 99US-0139457
PR      18-JUN-1999; 99US-0139458
PR      18-JUN-1999; 99US-0139459
PR      18-JUN-1999; 99US-0139460
PR      18-JUN-1999; 99US-0139461
PR      18-JUN-1999; 99US-0139462
PR      18-JUN-1999; 99US-0139463
PR      18-JUN-1999; 99US-0139750
PR      18-JUN-1999; 99US-0139763
PR      21-JUN-1999; 99US-0139817
PR      22-JUN-1999; 99US-0139899
PR      23-JUN-1999; 99US-0140353
PR      24-JUN-1999; 99US-0140354
PR      24-JUN-1999; 99US-0140695
PR      28-JUN-1999; 99US-0140823
PR      29-JUN-1999; 99US-0140991
PR      30-JUN-1999; 99US-0141287
PR      01-JUL-1999; 99US-0141842
PR      01-JUL-1999; 99US-0142154
PR      02-JUL-1999; 99US-0142055
PR      06-JUL-1999; 99US-0142390
PR      08-JUL-1999; 99US-0142803
PR      09-JUL-1999; 99US-0142920
PR      12-JUL-1999; 99US-0142977
PR      13-JUL-1999; 99US-0143542
PR      14-JUL-1999; 99US-0143624
PR      15-JUL-1999; 99US-0144005
PR      16-JUL-1999; 99US-0144085
PR      16-JUL-1999; 99US-0144086
PR      19-JUL-1999; 99US-0144325
PR      19-JUL-1999; 99US-0144331
PR      19-JUL-1999; 99US-0144332
PR      19-JUL-1999; 99US-0144333
PR      19-JUL-1999; 99US-0144334
PR      19-JUL-1999; 99US-0144335
PR      20-JUL-1999; 99US-0144335
PR      20-JUL-1999; 99US-0144352
PR      20-JUL-1999; 99US-0144884
PR      21-JUL-1999; 99US-0144814
PR      21-JUL-1999; 99US-0145086
PR      21-JUL-1999; 99US-0145088
PR      22-JUL-1999; 99US-0145085
PR      22-JUL-1999; 99US-0145087
PR      22-JUL-1999; 99US-0145089
PR      22-JUL-1999; 99US-0145192
PR      23-JUL-1999; 99US-0145145
PR      23-JUL-1999; 99US-0145218
PR      23-JUL-1999; 99US-0145224
PR      26-JUL-1999; 99US-0145276
PR      27-JUL-1999; 99US-0145913
PR      27-JUL-1999; 99US-0145918
PR      27-JUL-1999; 99US-0145919
PR      28-JUL-1999; 99US-0145951
PR      02-AUG-1999; 99US-0146386
PR      02-AUG-1999; 99US-0146486

```

[illegible]

GenCore version 5.1.6
Copyright (c) 1994 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: Jun 24, 2003, 10:14:56 ; Search time 14.848s Seconds
(without alignments)
1062 435 Million cell updates/sec

Title: US-09-830-751-4
Perfect score: 2625
Sequence: 1 MSAATQAVPAPNOPEVFC YSLGAYTRKFTVTKVQKN SQN

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents-AA.*

- 1: /cgn2_6/prodata/1/1aa/5A.COMB.pep.*
- 2: /cgn2_6/prodata/1/1aa/5R.COMB.pep.*
- 3: /cgn2_6/prodata/1/1aa/6A.COMB.pep.*
- 4: /cgn2_6/prodata/1/1aa/6B.COMB.pep.*
- 5: /cgn2_6/prodata/2/1aa/FCUS.COMB.pep.*
- 6: /cgn2_6/prodata/1/1aa/backfiles.pep.*

pred. No is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1746	66	5	US-09-221-294-2	Sequence 2, Appli
2	930	35	8	US-09-134-001C-4383	Sequence 4383, Ap
3	930	35	4	US-09-655-270A-9	Sequence 9, Appli
4	926	35	3	US-09-651-941-9	Sequence 9, Appli
5	926	35	3	US-09-955-597-9	Sequence 9, Appli
6	902	34	4	US-08-513-841-2	Sequence 2, Appli
7	902	34	4	US-08-606-834-2	Sequence 2, Appli
8	902	34	4	US-08-942-673-2	Sequence 2, Appli
9	902	34	4	US-08-118-317-2	Sequence 2, Appli
10	884	33	7	US-08-134-001C-4541	Sequence 4541, Ap
11	818.5	31.2	4	US-09-351-224E-5	Sequence 5, Appli
12	753	28.7	4	US-09-134-001C-4346	Sequence 5, Appli
13	681	25.9	4	US-09-134-001C-4388	Sequence 4388, Ap
14	652	24.8	5	US-09-134-001C-4451	Sequence 4451, Ap
15	602	22.9	4	US-09-155-183-4	Sequence 4, Appli
16	483	15.4	13	US-08-346-611-2	Sequence 2, Appli
17	403	15.4	13	US-08-784-494-2	Sequence 2, Appli
18	394	15.0	4	US-09-134-001C-4701	Sequence 4701, Ap
19	191	7.3	5	US-08-962-061-2	Sequence 2, Appli
20	119.5	4.6	8	US-08-867-129-5	Sequence 2, Appli
21	101.5	3.9	8	US-09-134-001C-5438	Sequence 5438, Ap
22	95.5	3.6	5	US-08-467-822-31	Sequence 31, Appl
23	95.5	3.6	5	US-08-432-697-31	Sequence 31, Appl
24	95.5	3.6	5	US-08-466-248-31	Sequence 31, Appl
25	95	3.6	7	US-09-134-001C-4908	Sequence 4908, Ap
26	93.5	3.6	4	US-09-329-878-2	Sequence 2, Appli
27	93.5	3.6	3	US-09-105-537-33	Sequence 33, Appli

28	93.5	3.6	11877	4	US-09-105-537-6	Sequence 6, Appli
29	93	3.5	535	3	US-08-813-574-2	Sequence 2, Appli
30	92.5	3.5	4551	3	US-09-320-878-1	Sequence 1, Appli
31	92.5	3.5	4613	4	US-09-105-537-31	Sequence 31, Appli
32	92	3.5	662	3	US-09-058-489-15	Sequence 15, Appli
33	92	3.5	662	3	US-09-058-489-15	Sequence 15, Appli
34	91.5	3.5	548	2	US-08-467-822-32	Sequence 32, Appli
35	91.5	3.5	548	2	US-08-467-822-32	Sequence 32, Appli
36	91.5	3.5	548	4	US-08-432-697-32	Sequence 32, Appli
37	91.5	3.5	548	4	US-08-466-248-32	Sequence 32, Appli
38	90.5	3.4	382	1	US-08-186-813-2	Sequence 2, Appli
39	90	3.4	563	4	US-09-431-470-4	Sequence 4, Appli
40	89.5	3.4	443	6	5310667-2	Patent No. 5310667
41	89.5	3.4	444	1	US-08-476-008-53	Sequence 53, Appli
42	89.5	3.4	444	1	US-08-476-008-54	Sequence 53, Appli
43	89.5	3.4	444	1	US-08-306-063-53	Sequence 53, Appli
44	89.5	3.4	444	1	US-08-306-063-54	Sequence 54, Appli
45	89.5	3.4	444	1	US-08-833-485-53	Sequence 53, Appli

ALIGNMENTS

RESULT 1
US-09-221-294-2
Sequence 2, Application US/09221294
Patent No. 6268138
GENERAL INFORMATION:
APPLICANT: Riccardo Dalla-Favera and
Alessandro Massimo Gianni
TITLE OF INVENTION: A Retroviral Vector Capable of Transducing the
Adipocyte Proliferation Gene and Uses of Said
TITLE OF INVENTION: Adipocyte Proliferation Gene and Uses of Said
TITLE OF INVENTION: Vector
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM 330 466 DX2
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1 24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,294
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/BOOKET NUMBER: 42990-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO. 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 521 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-221-294-2

Query Match
Best Local Similarity: 66.5%, Score 1746, DB 4, Length 521,
Matches 333, Conservative 69, Mismatches 98,
Query 1 MSAATQAVPAPNOPEVFCYSLGAYTRKFTVTKVQKN SQN
DB 17 MSAATQAVPAPNOPEVFCYSLGAYTRKFTVTKVQKN SQN
Query 61 DKRGRGCAEQUSPWRMASHGKLNHLALTEK LTLAALTEKNGKPVIS 118

```

Db 77 DKAVKAAQAAQICSPWRTMDASGRGLLYKLADLLERDRLLATMESMNGCKLYSNA 136
QY 119 YLVDLMVILKCLRYAGWACKY--HCKTIPICGDFESYTHPEFVSWGCGILIPWNEPLLMQ 176
Db 137 YLNDLAAIKTKILRYAGWACKYLGAGGRTIPICGDFESYTHPEFVSWGCGILIPWNEPLLMQ 195
QY 177 AWKLGPAALATGNVVMKVAEDTELTALYVANLIKKEAGPPGVNIVPGPGPTAGAAIASH 236
Db 197 IWKIGPALSNGNIVVWKPABQPTLTALHVASLIKKEAGPPGVNIVPGPGPTAGAAIASH 256
QY 237 EDVDKVAFTSIEIGKVIQVAAAGSNLKVITLGLGKSPNIIIMSDADMWAVEQAAHFAIF 296
Db 257 MDIDKVAFTGSTVGKLIKKEAGKSNLKVITLGLGKSPCIVLADADLNAVFAHGVF 316
QY 297 FNOQCQACASRTFVGEIITYFVFWVSFAKASKPVVWNPFSKTPGQGVVFTQKKY 356
Db 317 YHQQCQCIASRTFVEESIYDEVRBSVERAKKYIIGNPITPGVTOGPGQICKQYDKILD 376
QY 357 YINTGKQETAKLKGAGJIAALPGYFTQPTVEIGVAGGMIIAKEELTGPVMOILKFKITEE 416
Db 377 LIESKKEGAKLECGSPWCKNGYFVQPTVFSNVTDMMIAKEELTGPVMOILKFKSLDD 436
QY 417 VVGRANNSTYGLAAAVFTKDLKANYLSQALQAGTVVWNCYDFVGAQSPFGQYKMSGGR 476
Db 437 VIKRANNFTYGLSAGVFTKDLKAITISSALQAGTVVWNCYGSWGAQCPFGGKMSGGR 496
QY 477 ELGFGYGIQAYTEVKTIVTKVPOKN 500
Db 497 ELGEGHPHEVTEVKTIVTKISQKN 520

```

RESULT 2

US-09-134-001C-4383

; Sequence 4383, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; CURRENT APPLICATION NUMBER: US/09/134,001C

; PRIOR FILING DATE: 1998-08-13

; PRIOR FILING DATE: 1997-11-08

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 4383

; LENGTH: 506

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4383

Query Match

Best Local Similarity 44.8%; Score 949; DB 4; Length 506;

Matches 207; Conservative 75; Mismatches 186; Indels 10; Gaps 6;

```

QY 21 NQIFINNEHDAVSRKTEFTNPSTGEVICQVAEGKTEIVIKAREPPEAFQSGSPWPKM 80
Db 19 NQVIDGFWVSSNKNTRDIINPYNQETIFTVAEGTKEDVERAIIAARSEFDGE-WKSL 77
QY 81 DASHSGRLINKLAULERQRTYIALETGNKPKVIVSYVGLIMVLCIPIYAGWAKY 140
Db 78 TSEVKKKKVRAVADKIKFNPEFLAKIETHDCKTIEFSY-ADMDIHNHVMYFAGLAIKD 136
QY 141 HGKTI--PIDGDFESYTRIEHPGVCCQILIPWNEPLLMQAWKLGPAALATGNVVMKVAFO 198
Db 137 GGEIINSPIP-NAESVWKEPVQVYTOITPWNYPILQASWKIATLATGCSLWKPSEIT 195
QY 199 PITAIYVANLIKAGPPGVNIVPGPGTGAAGAAIASHEDVIKVAFTSTGEIYGVIVAA 258
Db 196 PITTIVPELMFVFPKPGTINILVIGAGSEVGVWMS-HEFVIVSFVSTET-KHMKQA 255

```

```

QY 259 GSSNLKRVITLGLGKSPNIIIMSDADMWAVEQAAHFAIFENQAGVCAASKTFVQELAYLE 318
Db 256 -ANBVTIWALEEDENINTEFTAAQDFLAIVDAIINQGYHAGQVSA-SSEPLIVNCDK 314
QY 319 FVFSVARAKSPVWNPFSKTEPHQGVQDEGFKKILGYINISKGGZAKLLAGGGLA 376
Db 415 FEKALIDKYSKIKLGNFQDQTEMGIVISAURKLEGYMEVAKKQATATIALGKKPKPE 474
QY 376 -AUGGYFIQVTFGVDDQMTIAKEIFGVVQGLIKFKTIEFVVGRRANNSTYGLAAVET 444
Db 375 DLQAGLFFETVITDCTSMRIVQERFQGVVTFVGFALFEERAIKLANUSTYGLAAVET 444
QY 435 KILUKANYLSQALQAGTVVWNCYDFVGAQSPFGQYKMSGSGRELELYGLQAYLEKTV 492
Db 444 KQESKAGAPVANPEKILGVWNIINHPVAGAPWQSGVSGVPEIPEIPEIPEIPEIPEI 492

```

RESULT 3

US-09-655-270A-9

; Sequence 9, Application US/09655270A

; Patent No. 6329151

; GENERAL INFORMATION:

; APPLICANT: Rouvriere, Pierre E.

; TITLE OF INVENTION: High Density Sampling of Differentially Expressed Proteins

; FILE REFERENCE: R01011 US NA

; CURRENT APPLICATION NUMBER: US/09/655,270A

; PRIOR FILING DATE: 2000-09-05

; PRIOR FILING DATE: 60/120,702

; PRIOR FILING DATE: 1999-February 19

; PRIOR FILING DATE: 60/152,542

; PRIOR FILING DATE: 1999-September 04

; NUMBER OF SEQ ID NOS: 47

; SOFTWARE: MicroSoft Office 97

; SEQ ID NO 9

; LENGTH: 508

; TYPE: PRT

; ORGANISM: Rhodococcus erythropolis HL PM-1

US-09-655-270A-9

Query Match

Best Local Similarity 35.4%; Score 930; DB 4; Length 508;

Matches 194; Conservative 92; Mismatches 200; Indels 26; Gaps 4;

```

QY 9 VPAPNOQPEVFNQ----- 508
Db 1 LPTPSSPPHTCTDGAERKTRIVQALTSVVIVLQDITSSGLATLDSINPAAASHLA 60
QY 51 QVAPGLFQFVDPAPPDPPGAPALGSSWPPMMAASPSPTINPLADTFDPPPIYLALETLD 110
Db 61 SVAEATAAIVARAVEAAKAA--AKTQWHRMRAQTRIMPRYAALEERKTELADQSGRO 117
QY 111 NCKPPVVISVILGDMVKLEYYA-WACKYR-KIIP-EDGDFESYTRIEHPGVQGLIWN 170
Db 118 MKGPRESLGIDLPIMETLETFAGLVITKTEGRTTPAIRSRFLNYLLEPFLVAVATLWN 177
QY 171 FTLLMQAWKLGPAALATGNVVMKVAEQDELTALYVANLIKKEAGTGVNIVPGPGTGA 240
Db 178 FPAVJAVWKIATLATGCSLWKPSEITPWNYPILQASWKIATLATGCSLWKPSEIT 247
QY 231 AAIAASHEDVIKVAFTSTGEIYGVIVAAAGSSNIIKVTILEGKSPNIIIMSDADMWAVEQ 290
Db 236 NALVQHSVSKVIFEGSTFVWAGGQHPMA-ADPFLIASLEDEGSAVAFQSSKRAVAV 296
QY 291 ABHALFPNQCQYCCAGSKTFVQEDYIDFVWVSVAKASRVWGNIPFSKTEPAGQVDELO 350
Db 297 VQAMYSNQGCTCTAHSRLLEVRPIYDEKVELVQARVEAAVGDPLDPTETGLISAFQ 356
QY 351 FKKILGYINTGQEGAKLIGGASL-----AALRGCTFQVTFVDDQDMETIAKEIFGVIV 406
Db 457 FESVHSVYVSGTFEGALITSGGQSPGAGPGGYRPILESVIAUMBIARELEPGVLE 416
QY 497 QILKELIEFVWGPANN-LYGLAAVAETKILGVANYSQALQATVWVWNYVIVPQASQSE 496

```

Db 417 SVLPPEPREFRAITLANIVFRLAAGVITFQVGSALFFAQLTLAGNWNINSWGVLNAPSPY 476
QY 467 GGYKMSGSRFLCYQVQAYTEKVTIVKVPQ 498
Db 477 PGFGSGVGSNLCQAATESFTKEKSIWARLDZ 508

RESULT 4

US-09-651-941-9
; Sequence 9, Application US/09651941
; Patent No. 6355470
; GENERAL INFORMATION:
; APPLICANT: ROUVIER, PIERRE E
; APPLICANT: WALTERS, DANA M
; APPLICANT: RAINER, RUSS
; TITLE OF INVENTION: Genes Encoding Picric Acid Degradation
; FILE REFERENCE: BC1022 US NA
; CURRENT APPLICATION NUMBER: US/09/651,941
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,545
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 9
; LENGTH: 485
; TYPE: PRT
; ORGANISM: Rhodococcus erythropolis HL PM-1
US-09-651-941-9

Query Match 35.3%; Score 926; DB 4; Length 485;
Best Local Similarity 38.9%; Pred. No. 4 3e-90;
Matches 194; Conservative 91; Mismatches 194; Indels 20; Gaps 4;

QY 4 AATQAVAPNQPQVFNQIFINNEWHDAVSRTFTFVNPSIGFVTCQVAFKELVKA 63
Db 3 ALTSSVP-----LVIGDOLTPSSTGATFDSINPADGSHLASVAEATAADVARA 50
QY 64 REGPGFQOLGSPWRMDASHSGPLNPLADLIERPNTYLAALFTLDNGKPVVISYLDL 123
Db 51 VEAAKAA---ARTQWRPPAQRFLMEFYAALIERHKTELALQLOSRDMCKPIRESLGIDL 107
QY 124 DMVLKCLRYAGWADKTHGKTIPIDGFFSYIPHEFVGVVQIIPNPFLLMQANKLGPA 183
Db 108 PIMETLEYFAGLVTKIEGRTTPAGRFNLNLTREPVGVAITPWNFFAVQAVWKIAPA 167
QY 184 LATGNVVMKVAEOTPLTALYVANLKEAGFPQGVNIVPGFGPTAGAAIASHEDYDKVA 243
Db 168 LAMGNAIVLKFAQLAPLVPVALGELALEAGLPCLYNVLPGSGVACNALVQHPVSGKVT 227
QY 244 FTGSTELGRVIOVAAGSNLKRVTLELGGKSPNIIKSDADMWAVEQAHAFFNQGQCC 303
Db 228 FTGSTEVQQLGRMA AGFHTASLEHGKKSALVAFTGSSIFKAVAAVVFQAMYSNUGETC 286
QY 304 CAGSRTFVQEDIIYDEFVVRSVAKSVVGNPNPLFSTKEQAPVLETFKILGYINLKGQ 363
Db 287 TAPSKLIVREPIYDEVVELVQARVEAARVGPGLDPTETGRLISAEQRESVHSYVSGTE 346
QY 364 EGAKLLCGGII---AARDGVFIQPTVFGDVQDGMTIAKEFIIPGVMOILKFKTIEEVVG 419
Db 347 EGAILSSGQSPGSAFQGFYFPPLFSGVTALMPIARKEIFGPVLSVLPFEGEEAIT 406
QY 420 RANSTVGLAAAVFTKLDKANYLSQALQAGIIVWNCYGVFGAQSPFGYKMSGSGRELG 479
Db 407 LANDTVFLAAGVFTPNVGRALPFAQLTLDAGNWNINSWGVLNAPSPYRGFGSGVSDLG 466
QY 480 EYGLQAYTEKVTIVKVPQ 498
Db 467 QAAIESFTKEKSIWARLDZ 485

RESULT 5

US-09-955-597-9

; Sequence 9, Application US/09955597
; Patent No. 6461856
; GENERAL INFORMATION:
; APPLICANT: POUVIER, PIERRE E
; APPLICANT: WALTERS, DANA M
; APPLICANT: PALNER, RUSS
; TITLE OF INVENTION: Genes Encoding Picric Acid Degradation
; FILE REFERENCE: BC1022 US NA
; CURRENT APPLICATION NUMBER: US/09/955,597
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/152,545
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 9
; LENGTH: 485
; TYPE: PRT
; ORGANISM: Rhodococcus erythropolis HL PM-1
US-09-955-597-9

Query Match 35.3%; Score 926; DB 4; Length 485;
Best Local Similarity 38.9%; Pred. No. 3 3e-90;
Matches 194; Conservative 91; Mismatches 194; Indels 20; Gaps 4;

QY 4 AATQAVAPNQPQVFNQIFINNEWHDAVSRTFTFVNPSIGFVTCQVAFKELVKA 63
Db 3 ALTSSVP-----LVIGDOLTPSSTGATFDSINPADGSHLASVAEATAADVARA 50
QY 64 REGPGFQOLGSPWRMDASHSGPLNPLADLIERPNTYLAALFTLDNGKPVVISYLDL 123
Db 51 VEAAKAA---ARTQWRPPAQRFLMEFYAALIERHKTELALQLOSRDMCKPIRESLGIDL 107
QY 124 DMVLKCLRYAGWADKTHGKTIPIDGFFSYIPHEFVGVVQIIPNPFLLMQANKLGPA 183
Db 108 PIMETLEYFAGLVTKIEGRTTPAGRFNLNLTREPVGVAITPWNFFAVQAVWKIAPA 167
QY 184 LATGNVVMKVAEOTPLTALYVANLKEAGFPQGVNIVPGFGPTAGAAIASHEDYDKVA 243
Db 168 LAMGNAIVLKFAQLAPLVPVALGELALEAGLPCLYNVLPGSGVACNALVQHPVSGKVT 227
QY 244 FTGSTELGRVIOVAAGSNLKRVTLELGGKSPNIIKSDADMWAVEQAHAFFNQGQCC 303
Db 228 FTGSTEVQQLGRMA AGFHTASLEHGKKSALVAFTGSSIFKAVAAVVFQAMYSNUGETC 286
QY 304 CAGSRTFVQEDIIYDEFVVRSVAKSVVGNPNPLFSTKEQAPVLETFKILGYINLKGQ 363
Db 287 TAPSKLIVREPIYDEVVELVQARVEAARVGPGLDPTETGRLISAEQRESVHSYVSGTE 346
QY 364 EGAKLLCGGII---AALPFTYFIQPTVFGDVQDGMTIAKEFIIPGVMOILKFKTIEEVVG 419
Db 347 EGAILSSGQSPGSAFQGFYFPPLFSGVTALMPIARKEIFGPVLSVLPFEGEEAIT 406
QY 420 RANSTVGLAAAVFTKLDKANYLSQALQAGIIVWNCYGVFGAQSPFGYKMSGSGRELG 479
Db 407 LANDTVFLAAGVFTPNVGRALPFAQLTLDAGNWNINSWGVLNAPSPYRGFGSGVSDLG 466
QY 480 EYGLQAYTEKVTIVKVPQ 498
Db 467 QAAIESFTKEKSIWARLDZ 485

RESULT 6

US-08-513-841-2
; Sequence 2, Application US/08513841
; Patent No. 5753481
; GENERAL INFORMATION:
; APPLICANT: Niwa, Mineo
; APPLICANT: Saito, Yoshimasa
; APPLICANT: Ishii, Yoshinori
; APPLICANT: Yoshida, Masaru
; APPLICANT: Suzuki, Hiromi
; TITLE OF INVENTION: No. 5753481el L-sorbose Dehydrogenase and No. 5753481el L-s
; TITLE OF INVENTION: Dehydrogenase Obtained from Gluconobacter oxydans T-100

```

: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Oblon, Spivak, McClelland, Maier & Neustadt, P.C.
: STREET: 1755 Jefferson Davis Highway, Suite 400
: CITY: Arlington
: STATE: Virginia
: COUNTRY: USA
: ZIP: 22202
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: MS-PCs Editor
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/513,841
: FILING DATE: 01-NOV-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: DK 9304700-9
: FILING DATE: 08-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 241851/1993
: FILING DATE: 28-SEP-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: NORMAN F. OBLON
: REGISTRATION NUMBER: 24,618
: REFERENCE/DOCKET NUMBER: 18-909-0 PCT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-413-3000
: TELEFAX: 703-413-2220
: TELEX: 248855 OPAT UR
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 497 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: ORIGINAL SOURCE:
: ORGANISM: Gluconobacter oxydans
: STRAIN: T-100
: FEATURE:
: NAME/KEY: mat peptide
: LOCATION: 1..497
: IDENTIFICATION METHOD: experimentally
: US-08-513-841-2

```

```

Query Match 34.4%; Score 902; DB 1; Length 497;
Best Local Similarity 39.5%; Pred. No. 1.3e-87;
Matches 189; Conservative 98; Mismatches 186; Indels 6; Gaps 5;

QY 24 FINNEMHDAVSKKTPPTVNPSTGEVICQVAETGDKEDVTKAPRPPGAPQIASWPPRMVAS 84
Db 19 FIDGWR--AGKDFDRSSPAHDVPTWIPKCTREDLDEAVAAAPRAFENG--WAGLAAA 75
QY 84 HSGRLNRLADLRDRTYLALETLDNCKPYVISYLVLDLMDVLKCLRYAGWADKYHGK 143
Db 76 DRAAVLLKAAGLRRERDDIATWELNKKP--ISCAKGEIHHCTACPFMAAGAAAPMILHD 134
QY 144 TPIPDGD--FFSYTRHEPVGCGQIIPWNEPILMQAWKILGPALATCNVVMKVAEQTPITA 202
Db 135 TNNLGEGLGMVREPVGVLITPWNFPFILLCERAPFILLASGCTLVVKKPAEVTSAT 194
QY 203 LLYVANLIKACGPPKVVNIIVPCFGPTAGAAIASHTGVKVAFTGTHGIVQVAAQSSN 262
Db 195 LLIAPILADAGLPKGVFNVTGTGTGTGQAMTEHDIDMLSPGTSGTGVGKSCIHAAADSN 254
QY 263 LKPVITLLEAGKSPNIMSDADMWAVECAHEAFLENQCCGCCAGSPFTVGQEDYDERVVR 322
Db 255 LKKLGLGGKKNPVPVADSNLEDAADAVAGISFNTGCCVCCSSRLIIVERSVAEKFERL 314
QY 323 SVARAKSRVNPPIHNSKTFQGPQVDETQFKILGYINTGKQEGAKLLCGGGIA--ADRGYF 381
Db 315 VVPRKMEKTRVGDPFQETQIGAITTEAQNKTILDYIAGRAEGAKLLCGGGIVDFCKGY 374

```

```

QY 382 LGHTVEGCVQDPMTLAKTEIPTPVMDLILKFKTELEVVWVANNSTYGLAAAAVETKLLAKAN 441
Db 375 LGHTFTDKPSMGARQETPGVLASHEFDIVGATAIANETVYGLAASWMSKDIKRAL 434
QY 442 YLSQALQATVWNVYVWFJASDFGSDYMSGSPEDEYETQAVTEKTVIVVVFQGN 500
Db 435 AVIKAVKAKRQWNTIMASGELILDSIKQSWHDEADLYVVLVLTGDKAVHLLGSKS 494

```

```

RESULT 7
US-08-696-834-2
: Sequence 2, Application US/08696844
: Patent No. 5834263
: GENERAL INFORMATION:
: APPLICANT: Niwa, Mineo
: APPLICANT: Saito, Yoshimasa
: APPLICANT: Ishii, Yoshinori
: APPLICANT: Yoshida, Masaru
: APPLICANT: Hayashi, Hiromi
: TITLE OF INVENTION: Method for Production of Ketol-Gulonol Acid
: NUMBER OF SEQUENCES: 48
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Oblon, Spivak, McClelland, Maier & Neustadt,
: STREET: 1755 Jefferson Davis Highway, Suite 400
: CITY: Arlington
: STATE: Virginia
: COUNTRY: USA
: ZIP: 22202
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/696,834
: FILING DATE: 24-SEP-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 28612/1994
: FILING DATE: 25-FEB-1994
: ATTORNEY/AGENT INFORMATION:
: NAME:
: REGISTRATION NUMBER:
: REFERENCE/DOCKET NUMBER:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 413-3000
: TELEFAX: (703) 413-2220
: TELEX: 248855 OPAT UR
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 497 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: ORIGINAL SOURCE:
: ORGANISM: Gluconobacter oxydans
: STRAIN: T-100
: FEATURE:
: NAME/KEY: mat peptide
: LOCATION: 1..497
: IDENTIFICATION METHOD: experimentally
: US-08-696-834-2

```

```

Query Match 34.4%; Score 902; DB 2; Length 497;
Best Local Similarity 39.5%; Pred. No. 1.3e-87;
Matches 189; Conservative 98; Mismatches 186; Indels 6; Gaps 5;

QY 24 FINNEMHDAVSKKTPPTVNPSTGEVICQVAETGDKEDVTKAPRPPGAPQIASWPPRMVAS 84
Db 19 FIDGWR--AGKDFDRSSPAHDVPTWIPKCTREDLDEAVAAAPRAFENG--WAGLAAA 75

```

```

QY      84 HSGRLNPLADIFEDRTYLAALETHINCKPYYVLSYLVGLDMVLKCLPYVYAGWAKYHCK 143
DB      76 DPAAVI LKAAGLLPERRDDIAYWVIENGKP-ISOAKGEIDHCHACFEMAAGARMMLHGD 134
QY      144 TIPIDGG-FFSYTHPEPVGCGQITIPWNFTI LKQAMKFLGALATGNWVVKVAEQTEFLJA 202
DB      135 TNNLGEGLFGMWLRPEIGVGLITPWNFPFEMILCERAPFTIASGCTLVVKKPAFTSAT 194
QY      203 LYVANLKEAGFPVGVNIVPGPPTAGAAALASHEDYKVAFTGCTGTEIGRVIOVAGSSN 262
DB      195 LLLAEILADAGLPGKGVNVTGTGTGTGVTGOAMTEHQDIDMLSTGCTGVGKSCILHAADS 254
QY      263 LFFVTITIGKSPNTKMSDAMTWAVEQAHEALFFKGSJGZAGSKIEVLJEDYDEVVK 422
DB      255 LKKLGLGKGNPIVWFADSNLEDAADAVAFGTSFNTGQCCVSSSKLIVERSVAERFEL 314
QY      323 SVARAKSVGNPNDSKTEQGPQVDETFQFKKILGYINTGKQKAKILCGGGIA-ADRGYF 381
DB      315 VPKMEKIRVGPDPETQIGAITTEAQNRIILDYIAKKAEGAKLLCGGGIVDFGKGY 374
QY      382 IQPTVEGVDGDMTIAKEEIPGPMQILKTKTIEEVVGRANNSTYGLAAVFTKDLDRAN 441
DB      375 IQPTLFTDVKPSMGIAKDEIFGPVLASFHEDTVDDEAIAIANDTVYGLAASVWSKIDKAL 434
QY      442 YLSQALQAGTVMVNYVVFRAQSPFGKYSKSGSPGELGEYSGLAAYTEVKTVIVKVPQKN 500
DB      435 AVTRVRAGRFWNTINSGGPEITLGGFKSGWGRAGLYGVVEYTIKSVHIETCKRS 493

```

RESULT 8

US-08-942-673-2

Sequence 2, Application US/08942673

Patent No. 5861292

GENERAL INFORMATION

APPLICANT: Niwa, Mineo

APPLICANT: Saito, Yoshimasa

APPLICANT: Ishii, Yoshinori

APPLICANT: Yoshida, Masaru

APPLICANT: Suzuki, Hiromi

TITLE OF INVENTION: L-sorbose body hydrogenase and No. 5861292-1

TITLE OF INVENTION: L-sorbose body hydrogenase obtained from Glucobacter

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: Obion, Spivak, McClelland, Maier & Neustadt, P.C.

STREET: 1755 Jefferson Davis Highway, Suite 400

CITY: Arlington

STATE: Virginia

COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.56 inch, 1.44 Mb storage

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: MS-DOS Editor

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/942,673

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/513,841

FILING DATE: 01-NOV-1995

APPLICATION NUMBER: 08/930,700 q

FILING DATE: 08-MAR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 241871/1993

FILING DATE: 28-SEP-1993

ATTORNEY/AGENT INFORMATION:

NAME: NORMAN F. OBLON

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 18-909-0 PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-413-3000

```

TELEFAX: 703-413-4240
TELEX: 248855 OPAT OR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 497 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Glucobacter oxydans
STRAIN: T-100
FEATURE:
NAME/KEY: cat peptide
LOCATION: 1-497
IDENTIFICATION METHOD: experimentally
US-08-942-673-2

```

Query Match 34.4% Score 902; DB 2; Length 497;

Best Local Similarity 39.5%; Pred. No. 1, 3e-87;

Matches 189; Conservative 98; Mismatches 186; Indels 6; Gaps 5;

```

QY      24 PINNWHUAVSPKTFPTVNVSTGVICQVAFPGKKEIVFKAFESPFGAFQLGSPWRMDAS 83
DB      19 FIDRWP--AGKQCFEDFSSPAHEVPVTRIP--TEIDLEAFAAARAFENG--WAGLAA 75
QY      84 HSCPLNPLADIFEDRTYLAALFTIGNCKPYYVLSYLVGLDMVLKCLPYVYAGWAKYHCK 143
DB      76 DPAAVI LKAAGLLPERRDDIAYWVIENGKP-ISOAKGEIDHCHACFEMAAGARMMLHGD 134
QY      144 TIPIDGG-FFSYTHPEPVGCGQITIPWNFTI LKQAMKFLGALATGNWVVKVAEQTEFLJA 202
DB      135 TNNLGEGLFGMWLRPEIGVGLITPWNFPFEMILCERAPFTIASGCTLVVKKPAFTSAT 194
QY      203 LYVANLKEAGFPVGVNIVPGPPTAGAAALASHEDYKVAFTGCTGTEIGRVIOVAGSSN 262
DB      195 LLLAEILADAGLPGKGVNVTGTGTGTGVTGOAMTEHQDIDMLSTGCTGVGKSCILHAADS 254
QY      263 LFFVTITIGKSPNTKMSDAMTWAVEQAHEALFFKGSJGZAGSKIEVLJEDYDEVVK 422
DB      255 LKKLGLGKGNPIVWFADSNLEDAADAVAFGTSFNTGQCCVSSSKLIVERSVAERFEL 314
QY      323 SVARAKSVGNPNDSKTEQGPQVDETFQFKKILGYINTGKQKAKILCGGGIA-ADRGYF 381
DB      315 VPKMEKIRVGPDPETQIGAITTEAQNRIILDYIAKKAEGAKLLCGGGIVDFGKGY 374
QY      382 IQPTVEGVDGDMTIAKEEIPGPMQILKTKTIEEVVGRANNSTYGLAAVFTKDLDRAN 441
DB      375 IQPTLFTDVKPSMGIAKDEIFGPVLASFHEDTVDDEAIAIANDTVYGLAASVWSKIDKAL 434
QY      442 YLSQALQAGTVMVNYVVFRAQSPFGKYSKSGSPGELGEYSGLAAYTEVKTVIVKVPQKN 500
DB      435 AVTRVRAGRFWNTINSGGPEITLGGFKSGWGRAGLYGVVEYTIKSVHIETCKRS 493

```

RESULT 9

US-09-118-317-2

Sequence 2, Application US/09118317

Patent No. 6197562

GENERAL INFORMATION:

APPLICANT: Niwa, Mineo

APPLICANT: Saito, Yoshimasa

APPLICANT: Ishii, Yoshinori

APPLICANT: Yoshida, Masaru

APPLICANT: Suzuki, Hiromi

TITLE OF INVENTION: No. 6197562-1 L-sorbose Dehydrogenase and No. 6197562-2

TITLE OF INVENTION: L-sorbose Dehydrogenase obtained from Glucobacter

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: Obion, Spivak, McClelland, Maier & Neustadt, P.C.

STREET: 1755 Jefferson Davis Highway, Suite 400

CITY: Arlington

STATE: Virginia

; APPLICANT: Folkerts, Otto
; APPLICANT: Crasta, Oswald R.
; TITLE OF INVENTION: Compositions and Methods for Pathogenesis
; TITLE OF INVENTION: Detoxification
; FILE REFERENCE: 5718-111
; CURRENT APPLICATION NUMBER: US/09/351,224E
; CURRENT FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 487
; TYPE: PR1
; ORGANISM: Exophiala spinifera
US-09-351-224E 5

Query Match 31.2%, Score 818 5, DB 4, Length 487;
Best Local Similarity 40.0%, Pred. No. 1.1e-78;
Matches 195; Conservative 86; Mismatches 193; Indels 13; Gaps 11;
QY 16 PEVFCNQIFINNEHDAVSRIKFTVNP-SIGEVICQVAEDKEDVDKAREGRPGAFOLG 74
Db 5 PRFYKSEIFINNEFVSSKSGSEPLITINPDESTVATDVHANAADVSAVASQAVKKG 64
QY 75 SPWRMDASHGRLNRLADLIERDRTYLALETLDNGKPVISYLVLDLDMVZKCLRYA 134
Db 65 -PWKFTGAQRAACMLKPADLAENAEKLAPESTPTGPPVSMITHPTDIPNMSVPRYA 123
QY 135 GWADKYHGKTIPIIDGDFSYTRHEPVGWCGOILPWNFLLMQAWKLGCPALATGNVVMKV 194
Db 124 GWADKIASKTFFEDNGKNNPYPGMSVACAGTASNNATFIYVGMKTAAPALAAAGSFIFKA 142
QY 195 ABOTRLTALYVANLIKAEAGFPFGVGVNIVFGPTAGAAIASHEHVDKVAFTSTGTEGRVI 254
Db 183 SEKSLPLVGLAPLFAEAGFPVGVVPLTGARVT-GEALASHMDIAKISITRSVGGRAV 241
QY 255 QVAAGSSNLKRVTLGKGSNNIIMSDADMWAVEQAHPALFNG-QGCCACSRPVVQ- 312
Db 242 KQATIKSNMKRVTLGKGR-PTIVNEAPLE-PQSGERAKDFSKFQILWVPSGLIVJW 298
QY 313 EDIYDEF-VVPSVAPAKSPVVS NPLSKHEGCPVVEIGKFFIIGVINTGKPPAKIL 459
Db 299 GNAEKFBVRHSGSGGQRLWGQNPLEPKETHEFFVKSQYDKVLGNIDVGK-DIACLL 357
QY 370 CGGGIAAUGYFLOPTVGVVQGMIAKEELFGVPMJILKPKTIEVVGKANNSTYGLA 429
Db 358 TGVGRKGDKGAIEPTIEVNEKPGSKIWFEBIFGPNLSIKTFKTEEAIBIANDTTYGIA 417
QY 430 AAVFTKDLKANYLSQALQAGTVVWNYVPSAUSPPRNVKMSGSHPELSEYSLDAYTPV 474
Db 418 SVYTKSLNRLVRSSALETGGVSNFPPTPETPTPGMKMGSGSGRELGEGLKAYLEP 477
QY 490 KTVTKV 496
Db 478 KTIINH 484

RESULT 12
US-09-134-001C-4246
; Sequence 4246, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4246

; LENGTH: 488
; TYPE: PR1
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4246

Query Match 28.7%, Score 753; DB 4; Length 488;
Best Local Similarity 35.8%, Pred. No. 1.1e-71;
Matches 173; Conservative 96; Mismatches 178; Indels 34; Gaps 11;

QY 24 FINNHDAVSKTFTVNPSTGVICQVAKGKEDVIAKAPGPPGAFQUGSPWRMDAS 83
Db 21 VINGEWDSASGETIDVINPATEEVMGKIAGNEEDVKNKAVD--AADKVYLEFRSSVE 77
QY 84 HSGRLNPLADLIERDRTYLALETLDNGKPVISYLVLDLDMVZKCLPYAGWAPKYHGK 143
Db 78 ERPELIGIKVKEYQNPKNMCIFATIDGAPLSVSPNHYCM--GNHPTAAPD----- 129
QY 144 TIPTINGDFSYTR-----HRPVGVCQITPWNFPIIMQAWKLGCPALATGNVVMKVAEQ 197
Db 130 --ALDSFOFEQRGDDLVVKEAIGVAGLVTPWNPFTNOTSLKLAFAAAGSPVVLKPSSE 187
QY 198 TPLTALYVANLIKAEAGFPFGVGVNIVFGPTAGAAIASHEHVDKVAFTSTGTEGRVIQVA 257
Db 188 TPFMAIHLAEIFDKVGPKEGVENLVNGDGSVGNPLSEHPKVRMMSTGSGTCSKIMEK 247
QY 258 AGSSNLKRVTLGKGSNNIIMSDADMWAVEQAHPALFNGQGCCACSRPVVQEDIVD 317
Db 248 A-AKDFKVSLELGKSPYIVLDVDVFEAANATTKKVVNTGAVTAGTRVILPSTIKE 406
QY 318 FEVHVSAPAKSPV-VGNPFSKTEGQPVVFTGFKKTIIGYINTGKGFAPILVQSG---- 372
Db 307 DYLI-AVREATSKVKGQFPEESTLVGPIISKKQFLQVQGYILKGINESAELEYGSPGKP 365
QY 373 -GIAADRYFTOPTVFGVDQGMIAKEELFGVPMJILKPKTIEVVGKANNSTYGLAAA 431
Db 366 ESL--DKGYFAPRTIFINVDNHMTIAOBEITFGPVMSVITYNNLDEATEIANDTKYGLAY 423
QY 432 VETREKDKANYLSQALQAGTVVWNYVPSAUSPPRNVKMSGSHPELSEYSLDAYTPV 489
Db 424 VICKIKITLPHVASTIACTTIFIN--PAGKPTIDPQGVKFSGLQFWQVYGEELLEV 480
QY 490 KTV 492
Db 481 KSI 483

RESULT 13
US-09-134-001C-4246
; Sequence 4246, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4388
; TYPE: PR1
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4388

Query Match 25.9%, Score 681; DB 4; Length 493;
Best Local Similarity 33.1%, Pred. No. 5.6e-64;
Matches 159; Conservative 97; Mismatches 201; Indels 24; Gaps 7;
QY 17 EVF--CNJFINNEHDAVSRIKFTVNPSTGIVLQVAESKEDVIAKAPGPPGAFOLG 74

```

Db      16  EVFNTNQLFINNEFISOSKETMDVNPATGSAFIIITLATEEVNVAIAEKSAQA-QI- 73
QY      75  SPWRMCAASHGRILNMIALIEPORTYLAALITLNN-KPYVIVSYLVIMLVKSLPYFA 134
Db      74  -EMERVQPOTRAEHRKILLIPLEKRNDEIAQLYVKQKTLAQAY-GEIDKSLIS-IDYMI 131
QY     135  GWADKYHCK - ----TIPIDGDFESYTRHEPQVWQGLIPWNPFLIMQAWKISPALA 185
Db     132  SLSMSDKRVLQNSIANETIQI-----INPKIGVTAGIVPWNAPILVLMRKVIPAIIV 183
QY     186  TGNVVMKVAEOPPLTALYVANLIKEAGPPPGVNNIVPGPPTAGAAATASHEDVDKVAFT 245
Db     184  TGSVVIKPSFEITLLTLRLAELFRASITIPAGLFGIVPGTGETVGTQLASHKDIQLISIT 243
QY     246  GSTEIGRVIOVAGSSNIMKVTILEGCKSPNTIMSDADMDWAVEGAHFAIFNQGCCCA 305
Db     244  GSMRAGKSVYENA-AQTQKVNILEGKNAPVIVTNSADLDKAVNYIVTARINNAGQVCTC 302
QY     306  GSRTFVQEDIVDFEVVRSVARAKSRVVGPNPDSKTPGQPVVDTPQFKITLGYINTCKQFG 465
Db     303  PERIFVEDVIDDELNKVTSKMKSLVYGPFFENTGYGALINQKULDSHEKVQQAIRNG 462
QY     366  AKLIGCGRIAADRYFIQPTVFGVQVQGMTIAKEELIPGPMQIIEKTEIEVVGPNANST 425
Db     363  ATLMTGSHGQI KPHGEFYAPTVLDNRKQVNVKDEIFGDLVLAITTYRDFEQVIEDANDTN 422
QY     426  YGLAAAVFTKDLKANKYLSQAQGVVWVYDVFCAQSPFGGKYMKSQSGRELGEYGLQA 485
Db     423  AGUSSYIFSENLTVEVMTATERLKFGEVYANCEAEVNVNYHAGWRESGLGGADGTHGPEE 482
QY     486  Y 486
Db     483  Y 483

```

RESULT 14

```

US-09-134-001C:4451
; Sequence 4451, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Dourcotte-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 6P0-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR FILING DATE: 1997-11-08
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4451
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C:4451

```

```

Query Match      24.8%; Score 652; DB 4; Length 518;
Best Local Similarity 34.9%; Pred. No 7 6e-61;
Matches 169; Conservative 82; Mismatches 209; Indels 24; Gaps 15;
QY     23  IFINNEHJAVSRKTFETVNP-STGEVIGQVAREGKEDVDKAREGRPGAFQSGFWRMD 81
Db      43  LVINGE--KLTKTDDTNSVNPANTSQILAKVSKATQDDIETAKAFESANHAYOSWRKWSHKD 100
QY     82  ASHSGRLINKLADIERDRTYLAALLETLDNGKPPVIVSYLVLDLDMVKLCIRYA-GWADKY 140
Db     101  RA---ELLIVAAIIIRREEISAINVYEGAPW-DEAVGAAEGIDETIYAKSMELA 156
QY     141  HGK-TIPIDGDFESYTRHEPQVCGGQIIPWNPFL-LMQAWKILPALATGNVVMKVAEQT 198
Db     157  DCKPVLDRGEHNRYP-YKPIGTGVTIPWNPFPALMAGTTLAPVVA-GNTVLLKPARDT 214

```

```

QY     199  FLIALYVANLIKAGFPVGVVNNIVPGPPTAGAAATASHEDVDKVAEQLUSTEIG 254
Db      215  VLTAYLMELEEAFLGPPVWNVVPPRFPD-DYLVVHEDDDEIVFDSTRATIDYERS 274
QY     254  IYVAASSNKKRVITLEGKSPNIMSDALMDWAVEGAHFAIFNQGVVAGSKTFVQGE 414
Db     275  AVVFEQGLFLKPVIAHMGKQIALVIVNNVDLLAALAVISAFQSGKCSAIVADQ 434
QY     314  DIYDEFVVSVAAPKSPVWNPDSKTEGSPVDDETQFKKITGYINTCKQVAKTIDGSS 474
Db     335  DVHDEILEKALIQTKLTGNT-EENTFMGVINOKOPDKIKNYTETGCKKEG KLEIGGG 492
QY     374  JAAIKGYEIQPTVFGVQVQGMTIAKKEIDGIMUILLKPKITFEVVGKANNSTYGLAAVAF 444
Db     393  TLDSTGYFTPTFSGLSAIPIMQDEIEFGVVGFTKVKRDFDEATEVANDHYGLIGAVI 452
QY     434  TKDLKANKYLSQAQGVVWV -GYDVFGCAQSPFGGKYMKSQSGRELGEYGLQAYTEV 489
Db     453  TNHPEHWIKAVNFEEDVNIYINMGCTIAAVVGYHDPGCKMSSCHDAKIDSHDY LANELEG 511
QY     490  KTVT 493
Db     512  KVV5 515

```

RESULT 15

```

US-09-155-183-4
; Sequence 4, Application US/09155184
; Patent No. 6324011
; GENERAL INFORMATION:
; APPLICANT: Natbad, Atjan
; APPLICANT: Rhodes, Michael J.C.
; APPLICANT: Gasson, Michael J.
; APPLICANT: Walton, Nicholas J.
; TITLE OF INVENTION: PRODUCTION OF VANILLIN
; FILE REFERENCE: 20747/100
; CURRENT APPLICATION NUMBER: US/09/155,184
; CURRENT FILING DATE: 1999-05-03
; EARLIER APPLICATION NUMBER: PCT/GB97/00809
; EARLIER FILING DATE: 1997-03-24
; EARLIER APPLICATION NUMBER: GR96/06187
; EARLIER FILING DATE: 1996-03-23
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 4
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Pseudomonas fluorescens
US-09-155-183-4

```

```

Query Match      22.9%; Score 602; DB 4; Length 482;
Best Local Similarity 32.6%; Pred. No 1 5e-55;
Matches 153; Conservative 94; Mismatches 199; Indels 24; Gaps 11;
QY     36  KTEFTVNPSTGEVIGQVAREGKEDVDKAREGRPGAFQSGFWRMDAS HSGRLINKLAD 94
Db      19  IREFENPVTVGELVSPVAAATLEDAVAVAACQAF---PAAWALAPNERSKLIKAAG 75
QY     95  LIEPDTYLAALLETLDNGKPPVIVSYLVLDLDMVKLCIRYA-GWADKYHKTIPD GD 150
Db     176  LQARSGEITAA-----GETGAMWNYGNVVLAAANMLKEAASMTTVGVGVVTSVIGS 140
QY     151  PFSYTRHEPQVCGGQIIPWNPFL-LMQAWKILPALATGNVVMKVAEQTITALLYANLIR 210
Db     131  FAWALR-OPCGVVGIGIAPWNAIVVILATRALAMTACGNVTIVLKAESLSPAVIRLIGVIG 189
QY     211  EAGFPVAVVNVVNGFPTACAAIA-----SHEHIVKVAETGTEIDPVLVVAAGSSNKKRV 266
Db     190  DAGIJDGVVNVVTSN-APADAAGIVETLANTAVRPVNETSTTHVGRVIG FLSAHIKIA 247
QY     267  TLPIGCKSPNIMSDALMDWAVEGAHFAIFNQGVVAGSKTFVQGEIDYEFVVSVAR 426

```



```
Db 248 LLEUGKAPILLVIFQATLEAAVGAARFGAYFNQGTICMSTERLIVDAKVDAPVAGLAAR 307
QY 327 AKSRVVGNPDSKTEQGPQVIDETQFKKILGY INIGKGGAKLLCGGGIAADRGYFIQPTV 386
   :: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 308 VETTFACUPADPESVWASU VNASAGTF IKALIDUAVAKGARI VIGGQL ---EGSTIQPTL 364
QY 367 FGDVGGMTIAKEETIGGVWGLLKFTIEEVVGKANNSTYGLAAAVFTKDLKANYLSQA 446
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 365 LGGVARSRLYREESFGVAVVLNGSTEFALDGLANISRFPSAAIFSWTGPALALAP 424
QY 447 LQACTVWNVYDVFG-AUSPPGRYKMSGGRELGEYGLQAYIEVKIVTK 495
   :: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 425 VESGICHINGPTVHDEAUMPFEGGVKSSGGYGGFGGKASIEHFQLRWVTIQ 474
```

Search completed: June 24, 2003, 10:31:40
Job time : 15.8469 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.
OM protein - protein search, using sw model

Run on: June 24, 2003, 10:17:21 ; Search time: 24.1632 seconds
(without alignments)
2398.528 Million coll updates/sec

Title: US-09-830-751-4
Perfect score: 2625
Sequence: 1 MSAATGAVPAPNCPVEVFC... YGICAVIEVIVIKVYUKN 500

Scoring table:
BLOSUM62
Gapop 10 0 Gapext 0 5

Searched: 417779 seqs, 108206813 residues
Total number of hits satisfying chosen parameters: 417774

Minimum seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match: 0%
Listing first 45 summaries

Database: Published_Applications_AA*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/2/pubpaa/BCI_NEW_PUB pep.*
3: /cgn2_6/ptodata/2/pubpaa/MS06_NEW_PUB pep.*
4: /cgn2_6/ptodata/2/pubpaa/MS06_PUB pep.*
5: /cgn2_6/ptodata/2/pubpaa/MS07_NEW_PUB pep.*
6: /cgn2_6/ptodata/2/pubpaa/MS07_PUB pep.*
7: /cgn2_6/ptodata/2/pubpaa/MS07_PUB pep.*
8: /cgn2_6/ptodata/2/pubpaa/MS08_PUB pep.*
9: /cgn2_6/ptodata/2/pubpaa/MS09_PUB pep.*
10: /cgn2_6/ptodata/2/pubpaa/MS10_PUB pep.*
11: /cgn2_6/ptodata/2/pubpaa/MS10_PUB pep.*
12: /cgn2_6/ptodata/2/pubpaa/MS10_PUB pep.*
13: /cgn2_6/ptodata/2/pubpaa/MS10_PUB pep.*
14: /cgn2_6/ptodata/2/pubpaa/MS10_PUB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	1771	67.5	412	10	US-09-925-302-736
2	1778	65.4	512	9	US-09-961-403-12
3	1718	65.4	512	9	US-10-268-518-2
4	1718	65.4	512	9	US-10-205-823-14
5	1543	58.8	538	9	US-09-344-882-20
6	1543	58.8	538	9	US-10-293-855-20
7	1473	56.1	534	9	US-09-344-882-24
8	1473	56.1	534	9	US-10-293-855-24
9	1305.5	49.7	496	9	US-09-847-208-59
10	1280	48.8	501	9	US-09-344-882-22
11	1280	48.8	501	9	US-10-293-855-22
12	1254	46.3	495	9	US-09-847-208-11
13	1214	46.2	314	10	US-09-925-300-1348
14	1201.5	45.8	492	9	US-10-268-518-4
15	1191	45.4	493	9	US-10-175-696-21
16	1191	45.4	493	10	US-09-823-901-9
17	1128	43.0	500	9	US-10-166-087-4
18	978.5	37.3	490	16	US-09-815-24-1057
19	975.5	37.2	496	10	US-09-815-24-10550

20	947.5	36.1	490	10	US-09-815-24-12102	Sequence 12102, A
21	947.5	36.1	496	10	US-09-815-24-5644	Sequence 5644, Ap
22	947.5	36.1	496	10	US-09-815-24-12557	Sequence 12557, A
23	933.5	35.6	518	9	US-09-919-039-143	Sequence 143, App
24	917.5	35.0	508	1	US-08-781-986A-5241	Sequence 5241, Ap
25	914.5	34.8	506	9	US-09-738-626-6572	Sequence 6572, Ap
26	833	31.4	487	9	US-10-175-696-17	Sequence 17, App1
27	823	31.4	487	10	US-09-815-24-5905	Sequence 5, App11
28	818.5	31.2	487	9	US-09-847-208-55	Sequence 3588, Ap
29	759.5	28.9	490	9	US-09-815-24-10574	Sequence 10574, A
30	759	28.9	481	10	US-09-815-24-10574	Sequence 10574, A
31	756	28.8	481	10	US-09-815-24-10574	Sequence 10574, A
32	754	28.7	475	10	US-09-815-24-12375	Sequence 12375, A
33	753	28.7	482	10	US-09-815-24-10407	Sequence 14047, A
34	744	28.3	493	10	US-09-815-24-5058	Sequence 5058, Ap
35	695.5	26.5	487	9	US-09-738-626-4108	Sequence 4108, Ap
36	662.5	25.2	491	10	US-09-497-664-3	Sequence 3, App11
37	654.5	24.9	521	9	US-09-738-626-6482	Sequence 6482, Ap
38	630	24.0	498	9	US-10-272-419-12	Sequence 12, App1
39	629	24.0	284	10	US-09-925-302-736	Sequence 189, App
40	621.5	23.7	484	9	US-09-738-626-6430	Sequence 6430, App
41	606.5	23.1	481	1	US-08-976-0640-28	Sequence 28, App1
42	582	22.9	482	10	US-09-738-626-4108	Sequence 4108, Ap
43	548.5	20.9	479	10	US-09-815-24-1341b	Sequence 1341b, A
44	540	20.5	192	9	US-10-105-698-5298	Sequence 5298, Ap
45	522.5	20.3	453	9	US-09-738-626-4037	Sequence 4037, Ap

ALIGNMENTS

RESULT 1
US-09-925-302-736
Sequence 736, Applicant: US-09-925-302
Patent No. US-20020044941A1
GENERAL INFORMATION:
APPLICANT: Posen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: FA04
CURRENT APPLICATION NUMBER: US-09-925-302
CURRENT FILING DATE: 2001-08-10
PCT APPLICATION NUMBER: PCT/US00/05918
PCT FILING DATE: 2000-03-08
PCT APPLICATION NUMBER: 437124.275
PRIORITY FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 896
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 736
LENGTH: 412
TYPE: PRT
ORGANISM: Homo sapiens
US-09-925-302-736

Query Match 67.5% Score 1771; DB 10; Length 412;
Best Local Similarity 90.0% Freq. No. 6; 2e156;
Matches 393 Conservative 2; Mismatches 28; Indels 0; Gaps 0;
1 MSAATGAVPAPNCPVEVFC... YGICAVIEVIVIKVYUKN 500
|||||
61 FRAPDFAFCDSDSWFK-AASDSCGLNLAC-ILKGFYALDFTGQPVYISY 120
|||||
20 USAAAGAVAPNCPVEVFC... YGICAVIEVIVIKVYUKN 500
|||||
61 FRAPDFAFCDSDSWFK-AASDSCGLNLAC-ILKGFYALDFTGQPVYISY 120
|||||
80 FFVAVAVAVATGCTSWFPMOAFHPCETLNPLAV TFFPTVIALFTTCTGFVYISY 139
|||||
121 VGLGAVV-KLNYAYAWAKIKYKLTFTGQVSTSRFPVAVGQITPMNPIIMAWCI 180
|||||
140 VGLGAVV-KLNYAYAWAKIKYKLTFTGQVSTSRFPVAVGQITPMNPIIMAWCI 180
|||||
181 GATATGAVVAVKVAEOTPLATLVANLIKAKGPTPVVAVIYGFETATATASHEVY 240
|||||
430 GATATGAVVAVKVAEOTPLATLVANLIKAKGPTPVVAVIYGFETATATASHEVY 253


```

: APPLICANT: Monahan, John E.
: APPLICANT: Endege, Wilson O.
: APPLICANT: Gannavarapu, Manjula
: APPLICANT: Gorbacheva, Bella
: APPLICANT: Hoeisch, Sebastian
: APPLICANT: Kamatkar, Shubhang
: APPLICANT: Wensev, Angela M
: APPLICANT: Glatt, Karen
: APPLICANT: Zhao, Xumei
: APPLICANT: Anderson, Rustin
: TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
: TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
: TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
: FILE REFERENCE: MFI-044
: CURRENT APPLICATION NUMBER: US/10/205,823
: CURRENT FILING DATE: 2002-07-25
: PRIOR APPLICATION NUMBER: 60/307,982
: PRIOR FILING DATE: 2001-07-25
: PRIOR APPLICATION NUMBER: 60/414,456
: PRIOR FILING DATE: 2001-08-22
: PRIOR APPLICATION NUMBER: 60/325,020
: PRIOR FILING DATE: 2001-09-25
: PRIOR APPLICATION NUMBER: 60/341,746
: PRIOR FILING DATE: 2001-12-12
: PRIOR APPLICATION NUMBER: 60/362,158
: PRIOR FILING DATE: 2002-03-05
: NUMBER OF SEQ ID NOS: 455
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 14
: LENGTH: 512
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-205-823-14

```

```

Query Match 65.48, Score 1718, DR 2, Length 512.
Best Local Similarity 65.58, Pred. No 7 of 151;
Matches 323; Conservative 66; Mismatches 104; Indels 0; Gaps 0;

QY 8 AVFAPNQPEVFCNQIFINNEHUAUSKRIPTVNPSTGEVTCQVAREKDEVDKARGP 67
Db 19 ALPRIRNLVVKFKIFINNEHUESKSKKFAFCNSTREQICEVEEGDKPDVKARAA 78
QY 68 PGTQGLSPPPMPASHSGLLNPLADLIERDTYLAALLETLDNCKPYVSYLVLDVYL 127
Db 79 QVAFQSPSPRLDALSFGRLHLQDLADLVERDEATLAALLETMDTKPFLHAFIDEGCI 138
QY 128 KCLRYAGWADKYHGKTIPTDGGFFSYTRHPVGVGGQIIPNFPILMCAWKLPALATG 187
Db 139 RLYFPAQWADKTOGKTIPTDDNVVCFTRHPGLVCGAITPNFPILMLVWLAPALCG 198
QY 188 NVVMYKVAEOTPLTALYVNLKEAGFPVGVNIVPGFPTAGAAATASHEDYDKVFTGS 247
Db 199 NTMVLKPAEOTPLTALYVNLKEAGFPVGVNIVPGFPTVGAALISSHPQINKIAFTGS 258
QY 248 TEIGFVTCVAGSSNLKEVILEGGKSFNINSLAUMENWAVEAHEAFEFNGQCCCA'S 307
Db 259 TEVGLVKEAASNLKRVTELEGGKNCIVCAUADLDLAVECAHGVFEINQGCCTAAS 318
QY 308 RTHVQEDIDYEFVVRVARAKSVVNPFDUSKTEQGHPVDETFQFKILCYINTGKQEGAK 367
Db 319 RVFVEEQVYSEFVPRSVYAKRPVGDPEVDVTEQGPQIDQKQFLKLELIESGKKEGAK 378
QY 368 LIGRNIAAIOGYFQPTVPELWQJQMIIAKEEIFGVPMQILKFKTIEEVGRANNISYG 427
Db 379 LECGGSAMEDGLFKTPKTFSEVETDNNRIAKEEIFGVPMQILKFKSIEEVIKRANSTDYG 438
QY 428 LAAAVETKLDLKANYLSUALQAGTVVWVNCYDVFCAUSPPGQYKMSGSGRELGEYGLAYT 487
Db 439 LTAAVETKLDLKALIKASLESIGTVWVNCYNALYAOPFGGKMSGNGRELGEYALAYT 498
QY 488 EVKTVTVKVPQKN 500
Db 499 EVKTVTVKLGDNK 511

```

```

RESULT 5
US-09-344-882-20
: Sequence 20, Application US/09344882
: Patent No. US 20020162137A1
: GENERAL INFORMATION:
: APPLICANT: Nikolau, Basil J
: APPLICANT: Wurtele, Eve S
: APPLICANT: Oliver, David J
: APPLICANT: Behal, Robert
: APPLICANT: Schnable, Patrick S
: APPLICANT: Ke, Jinshan
: APPLICANT: Johnson, Jerry L
: APPLICANT: Allred, Carolyn C
: APPLICANT: Fatland, Beth
: APPLICANT: Lutziger, Isabelle
: APPLICANT: Wee, Tsui-Jung
: TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
: TITLE OF INVENTION: Acetyl CoA Levels in Plants
: FILE REFERENCE: 201573
: CURRENT APPLICATION NUMBER: US/09,344,882
: CURRENT FILING DATE: 1998-06-25
: PRIOR APPLICATION NUMBER: US 60/090,717
: PRIOR FILING DATE: 1998-06-26
: NUMBER OF SEQ ID NOS: 38
: SOFTWARE: PatentIn Ver. 2.2
: SEQ ID NO 20
: LENGTH: 538
: TYPE: PRT
: ORGANISM: Arabidopsis thaliana
US-09-344-882-20

```

```

Query Match 58.88, Score 1543, DB 9, Length 538;
Best Local Similarity 61.14, Pred. No. 1 of 134;
Matches 300; Conservative 61; Mismatches 124; Indels 4; Gaps 2;

QY 2 SAAATQAVPAPNQPEVFCNQIFINNEHDAVSRKTFPTVNPSTGEVTCQVAREKDEVD 61
Db 42 SAAAEII---NFSVQVSHQTQLLINFVDSASGKTFTTIDPTGEVIAHVAEGDAEDIN 98
QY 62 KAREGHPHAPQJASPMPPMASHSGLLNPLAAILPPIHPTYLAALLETLDNCKPYVSYLV 121
Db 99 PAVKAAFTATPTFG-PMPKMSAYERSVLLRFADLVEKHEEELASLETWDNCKPYQOOSLTA 157
QY 122 DLDWVLKCLRYAGWADKYHGKTIPTDGGFFSYTRHPVGVGGQIIPNFPILMCAWKIG 181
Db 158 EIPWFAPLFPYYASNAIKTHGLTIPAGGNVYVHTLHEPISVAGSQIIPNFPILMFAHVG 217
QY 182 PALATGNVVMYKVAEOTPLTALYVNLKEAGFPVGVNIVPGFPTAGAAATASHEDYDK 241
Db 218 PALACGNTIVIKTAFETPTAFYAGKLFLEAGLPIYVIVSGFATAZALASHMUVDK 277
QY 242 VATGSTEIGRPVTCVAGSSNLKEVILEGGKSFNINSLAUMENWAVEAHEAFEFNGQ 301
Db 278 LAFETSTUTSKVILDLAANSNKPVTFIEGRSPFIVFEADITKRAVETIAHEAFEFNGQ 337
QY 302 CCAGSPTVPCPIYHFFVVRVARAKSVVNPFDUSKTEQGHPVDETFQFKILCYINTG 361
Db 448 CCAGSPTVPCPIYHFFVVRVARAKSVVNPFDUSKTEQGHPVDETFQFKILCYINTG 397
QY 362 KQEVAKLIGRNIAAIOGYFQPTVPELWQJQMIIAKEEIFGVPMQILKFKTIEEVGR 421
Db 398 IESNATLECGDGLTGKGYFQPTVPELWQJQMIIAKEEIFGVPMQILKFKSIEEVIKRA 457
QY 422 NNSTYGLAAVETKLDLKANYLSUALQAGTVVWVNCYDVFCAUSPPGQYKMSGSGRELGEY 481
Db 458 NERYGLAAVETKLDLKANYLSUALQAGTVVWVNCYDVFCAUSPPGQYKMSGSGRELGEY 517
QY 482 GLQAYTEVTKV 492
Db 518 SLNNYLOIKAV 528

```

RESULT 6

US-10-293-865-20
 : Sequence 20, Application US/10293865
 : Publication No. US20020106090A1
 : GENERAL INFORMATION:
 : APPLICANT: Nikolau, Basil J
 : APPLICANT: Wurtelo, Eve S
 : APPLICANT: Oliver, David J
 : APPLICANT: Behal, Robert
 : APPLICANT: Schnable, Patrick S
 : APPLICANT: Ke, Jinshan
 : APPLICANT: Johnson, Jerry L
 : APPLICANT: Allred, Carolyn C
 : APPLICANT: Falland, Beth
 : APPLICANT: Lutziger, Isabelle
 : APPLICANT: Wen, Tsui-Jung
 : TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
 : FILE REFERENCE: 217113
 : CURRENT APPLICATION NUMBER: US/10/293,865
 : CURRENT FILING DATE: 2002-11-13
 : PRIOR FILING DATE: 1999-06-25
 : PRIOR APPLICATION NUMBER: US 09/344,882
 : PRIOR FILING DATE: 1998-06-26
 : NUMBER OF SEQ ID NOS: 38
 : SOFTWARE: PatentIn Ver. 3.1
 : SEQ ID NO 20
 : LENGTH: 538
 : TYPE: PRT
 : ORGANISM: Arabidopsis Thaliana
 US-10-293-365-20

Query Match 58.8% Score 1543; DB 9; Length 538;
 Best Local Similarity 61.1% Pred. No. 1,5e-134;
 Matches 300; Conservative 61; Mismatches 126; Indels 4; Gaps 2;

QY	2	SAAATQAVPAPNQOFEVFCNQIFINNEWHDAVSRTFTVPNPSTGEVLCQVAESGDKEDV	61
DB	42	SAAAEELI--NP-SVQVSHQTQLLINGNEFVDSASGKTFTPLDPTGEVIAHVAEGDAEDIN	98
QY	62	KAREGRPCAFQASGPPMDASHSGRLNPLADLIERIKTYLAALETLNDCKPYPVSYLV	121
DB	99	RAVKAATAFDEG-PMPKMSAYERSRVLLRFADLVEKHSSEELASLETWDNGKPYQOOLTA	157
QY	122	DLDMVLKCLRYAGWADKYHGKTIPTDGDFFSYTRHEPVGCGQITPWNFPLLMQAWKLG	181
DB	158	EIPMFAHLPYYAGWAKIHGLTIPADGNYQVHTLHEPIGVAGQIIPWNFPLLMFAKVG	217
QY	182	PALATGNVVMKVAHQTPUTALYVANLKEAGFPFGVNVNIVPGFGPTAGAAIASHEDVDK	241
DB	218	PALACGNTIVLKTAEQTPITAFYAGKLFLEAGLPPGVLNIVSGFGATAGALASHMDVDK	277
QY	242	VAFGTSGTEIGRVIOVAGSSNLKRVTLLEGGKSPNIMSDALMDMWAVEQAHFALFNQGT	301
DB	278	LATGSDTGKVLGLAANSNLKPVTLLEGGKSPFIVFEDADIDKAVELAHFALFNQGT	337
QY	302	CCAGSRTFVOEDIDYDEFVVRVARAKSRVGNPFDSKTEGSPQVDETQFKKILGYINTG	361
DB	338	CCAGSRTFVHEKRVYDEFVEKSKAKALRVGDUPEFKGTQKQIQIDLKQFEKVMKYTKSG	397
QY	362	KQEKAKILCGGRTIAADPGYFIQPTVPFVQVQMTAKFEIFCPVMQILKFKTIFEVVGR	421
DB	398	IESNATLEGGDQIDGKGFYFIQPTVPFVNSVDMLAQDEIFGPVUSILKFSDDVDVFKRA	457
QY	422	NNSTYGLAAAVFTIKDLDKANYLSQALQAGTVWVNCYDFVGAQSPFGGKMSGSGRELGEY	481
DB	458	NETKYGLAACVFTKNLDTANRVSRALKAGTVWVNCYDFVGAQSPFGGKMSGSGRELGEY	517
QY	482	GLOAYTEVKTV	492
DB	518	SLNNYIQIRAV	528

RESULT 7

US-09-344-882-24
 : Sequence 24, Application US/09344882
 : Patent No. US20020162137A1
 : GENERAL INFORMATION:
 : APPLICANT: Nikolau, Basil J
 : APPLICANT: Wurtelo, Eve S
 : APPLICANT: Oliver, David J
 : APPLICANT: Behal, Robert
 : APPLICANT: Schnable, Patrick S
 : APPLICANT: Ke, Jinshan
 : APPLICANT: Johnson, Jerry L
 : APPLICANT: Allred, Carolyn C
 : APPLICANT: Falland, Beth
 : APPLICANT: Lutziger, Isabelle
 : APPLICANT: Wen, Tsui-Jung
 : TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
 : FILE REFERENCE: 201573
 : CURRENT APPLICATION NUMBER: US/09/344,882
 : CURRENT FILING DATE: 1999-06-25
 : PRIOR APPLICATION NUMBER: US 60/090,717
 : PRIOR FILING DATE: 1998-06-26
 : NUMBER OF SEQ ID NOS: 38
 : SOFTWARE: PatentIn Ver. 2.2
 : SEQ ID NO 24
 : LENGTH: 534
 : TYPE: PRT
 : ORGANISM: Arabidopsis Thaliana
 US-09-344-882-24

Query Match 56.1% Score 1473; DB 9; Length 534;
 Best Local Similarity 59.3% Pred. No. 4.5e-128;
 Matches 294; Conservative 55; Mismatches 143; Indels 4; Gaps 2;

QY	1	MSAATQAVPAPNQOFEVFCNQIFINNEWHDAVSRTFTVPNPSTGEVLCQVAESGDKEDV	60
DB	37	LA-AAVENTITTP--VKVEHTQLLGGREVDVAVSKTEPTLDPRTGEVIAHVAEGDAEDV	94
QY	61	DKAREGRPCAFQASGPPMDASHSGRLNPLADLIERIKTYLAALETLNDCKPYPVSYLV	120
DB	94	NNAAVAAAKAFDEG-PMPKMTAYERSKILFEPADLIERKNDLAALETWINDCKYEQSAQ	152
QY	121	VBLDMVLKCLRYAGWADKYHGKTIPTDGDFFSYTRHEPVGCGQITPWNFPLLMQAWKLG	180
DB	153	TEVPMFAHLPYYAGWAKIHGLTIPADGNYQVHTLHEPIGVAGQIIPWNFPLLMQAWKLG	212
QY	181	GPALATGNVVMKVAHQTPUTALYVANLKEAGFPFGVNVNIVPGFGPTAGAAIASHEDVDK	240
DB	213	GPALACGNTIVLKTAEQTPITAFYAGKLFLEAGLPPGVLNIVSGFGATAGALASHMDVDK	272
QY	241	KVAFGTSGTEIGRVIOVAGSSNLKRVTLLEGGKSPNIMSDALMDMWAVEQAHFALFNQGT	300
DB	274	KVAFGTSGTEIGRVIOVAGSSNLKRVTLLEGGKSPNIMSDALMDMWAVEQAHFALFNQGT	332
QY	301	CCAGSRTFVOEDIDYDEFVVRVARAKSRVGNPFDSKTEGSPQVDETQFKKILGYINTG	360
DB	333	CCAGSRTFVHEKRVYDEFVEKSKAKALRVGDUPEFKGTQKQIQIDLKQFEKVMKYTKSG	392
QY	361	KQEKAKILCGGRTIAADPGYFIQPTVPFVQVQMTAKFEIFCPVMQILKFKTIFEVVGR	420
DB	393	GVEGATILQAGGRISGKGYIQTPTVPFVNSVDMLAQDEIFGPVQVTLKFKELDEHVAR	472
QY	421	NNSTYGLAAAVFTIKDLDKANYLSQALQAGTVWVNCYDFVGAQSPFGGKMSGSGRELGEY	480
DB	453	NNSTYGLAAAVFTIKDLDKANYLSQALQAGTVWVNCYDFVGAQSPFGGKMSGSGRELGEY	512
QY	481	GLOAYTEVKTV	496
DB	513	YSLNYYIQIRAV	528

RESULT 8

US-10-293-865-24
 : Sequence 24, Application US/1029865
 : Publication No. US20030082190A1
 : GENERAL INFORMATION:
 : APPLICANT: Nikolau, Basil J
 : APPLICANT: Warteale, Eve S
 : APPLICANT: Oliver, David J
 : APPLICANT: Behal, Robert
 : APPLICANT: Schnable, Patrick S
 : APPLICANT: Ke, Jinsan
 : APPLICANT: Johnson, Jerry L
 : APPLICANT: Fatland, Ruth
 : APPLICANT: Lutziger, Isabelle
 : APPLICANT: Wen, Tsui Jung
 : TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
 : FILE OF INVENTION: Acetyl CoA Levels in Plants
 : FILE REFERENCE: 21713
 : CURRENT APPLICATION NUMBER: US/10/293,865
 : CURRENT FILING DATE: 2002-11-13
 : PRIOR FILING DATE: 1999-06-25
 : PRIOR APPLICATION NUMBER: US 60/090,717
 : PRIOR FILING DATE: 1998-06-24
 : NUMBER OF SEQ ID NOS: 38
 : SOFTWARE: PatentIn Ver. 3.1
 : SEQ ID NO: 4
 : LENGTH: 534
 : TYPE: PRT
 : ORGANISM: Arabidopsis Thaliana
 US-10-293-855-24

Query Match 56.18; Score 1473; DB 9; Length 534;

Best Local Similarity 59.38; Pred. No. 4.5e-124;

Matches 294; Conservative 55; Mismatches 143; Indels 4; Gaps 2;

QY 1 MSAAATQAPAPNOQPEVFCNQIFINNEHDAVSRKFTPTVNSTGEVQCVAEGKEDV 60
 DB 37 LAAAVENITTP---VKVEHTQLLIGRFVDVAVSGKFTPLDPRNGEIVAQVSEGDADV 93
 QY 61 KRAEPPGPGATQISPPWPMASHGRLINPLADIEPDPTYLAAETLNGKPYVLSYL 120
 DB 94 NRAVAAPAKADEG-PWPKMTAYERSKILFRFADLIEKHNDLAAETLNGKPYVLSYL 152
 QY 121 VLLMLVLLKLYAA-SWAL-KYHRSKTFIPFGDFFSVIPHEPVGVSQGLIPWNPFLMGAWKL 180
 DB 153 LEVPMALVPFYYAGWAKIHWMTMPGDPHHVOTLHEPIGVACQGLIPWNPFLMLSWKL 212
 QY 181 GPALATGNVVMKVAEOTPLTALYVANLIKAGPPTGVNIVPGPPTAGAAIASHEDVD 240
 DB 213 GPALACGNTVVLKTAEOITPLSALLVGLLHEAGLFDGWNIVSGGATAGAAIASMDVD 272
 QY 241 KVAFTGSTEIGRVIOVAGSSNLKRVTLLELGGKSPNLIIMSDADMWAVEQAHPALFNQ 300
 DB 273 KVAFTGSTIDVGKILILEASKSNLRAVTLLEESHHSEVCEADVQAVELAHFALFNQ 332
 QY 301 QCCAGSPFTFQVEDIYDFVVRVARAKSRVVGPNPDSKTEQSPQVDOTQKILGYINT 360
 DB 333 QCCAGSPFTFVHVWYDFVEVEKAKAPALKRVNGDPPKSGIEQSPQVDSEOPNKLKIKH 392
 QY 361 KQCEAKLKGSSGTAALPGYFIQPIVFTLWQMTIAKEELPGVLMILKPKILEEVVGR 420
 DB 393 GVEACATLQAGSDRUGSKGYIQPTVFSQKDDMLIATDEIEGVPQTILKFKDLDEVIAR 452
 QY 421 ANNSTYGLAAVFTKDLKKNYLSQALGAGIIVWVNYVDFCAQSPFGYKMSGSCKELCE 480
 DB 453 ANNRYGLAAGVFTQNLDTAHLKRALRVGTIVNCFDVIDASTIPFGYKMSGTPGPKGI 512
 QY 481 YGLQAYTEVKTIVTKV 496
 DB 513 YSLNNYQVKAIVTSL 528

RESULT 9

US-09-847-208-59
 : Sequence 59, Application US/09847208
 : Publication No. US20030082190A1
 : GENERAL INFORMATION:
 : APPLICANT: Saxon, Andrew
 : APPLICANT: Zhang, Ke
 : APPLICANT: Zhu, Daocheng
 : TITLE OF INVENTION: FUSION MOLECULES AND TREATMENT OF
 : FILE OF INVENTION: IGF-MEDIATED ALLERGIC DISEASES
 : FILE REFERENCE: UC67,002A
 : CURRENT APPLICATION NUMBER: US/09/847,208
 : CURRENT FILING DATE: 2001-05-01
 : NUMBER OF SEQ ID NOS: 177
 : SOFTWARE: FastSeq for Windows Version 4.0
 : SEQ ID NO 59
 : LENGTH: 496
 : TYPE: PRT
 : ORGANISM: Cladosporium herbarum
 US-09-847-208-59

Query Match 49.78; Score 1305.5; DB 9; Length 496;

Best Local Similarity 55.48; Pred. No. 1.4e-112;

Matches 262; Conservative 63; Mismatches 146; Indels 3; Gaps 3;

QY 23 IFINNEHDAVSPKFTPTVNSTGEVQCVAEGKEDVTKAPETKFAFLDLSIPWPMCA 82
 DB 20 LFINNEFVKGQCKTFDVINPDSVITQVHEATEKDVDAVAAAQAPE-GS-WRLETP 77
 QY 83 SHSGRLNRLADLIERDPTYLAAETLNGKPYVLSYVLDLWVLCIPIYYAGWADKYHG 142
 DB 78 ENRGKLLNLANLFKANTDLLAAVESLDNGKATSMARVISA-CASGCLPYYGWDKKTG 136
 QY 143 KTIPIGDFEFSTHREPVGCGQIIPWNPFLMGAWKLGPALATGNVVMKVAEOTPLTA 202
 DB 137 KVIDTTPDTFNKYVKKEPIGCRSDHSLPLLLMAMWIKIGPALACGNTVVLKTAEOITP 196
 QY 203 LYVANLIKAGPPTGVNIVPGPPTAGAAIASHEDVDKVAFTGSTEIGRVIOVAGSSN 262
 DB 197 LYAASLVKAGPPTGVNIVSGFGKAGAAALSHMDVDKVAFTGSTIVGRTILKAAASN 256
 QY 263 LKRVHLEGGSPFNLIIMSDADMWAVEQAHPALFNQSPGAGSPFTFQVEDIYDFVVP 322
 DB 257 LKRVHLEGGSPFNLIIMSDADMWAVEQAHPALFNQSPGAGSPFTFQVEDIYDFVVP 322
 QY 323 SVAPAKSPVVPNPFESKTEQSPQVDOTQKILGYINTQKPEAKILLGSGIAAUGYFI 482
 DB 317 FKPAKKNVVGSGPFAALPGYFIQPIVFTLWQMTIAKEELPGVLMILKPKILEEVVGR 476
 QY 383 QPTVFTGVQDGMTIAKEELPGVLMILKPKILEEVVGRANNSTYGLAAVFTKDLKANY 442
 DB 377 EPTFNSVTDKMTVKEEIPGVCSIAKPKTKEDAIKIGNASTYGLAAVHTKINTALE 436
 QY 443 LSQALGAGTVWVNYVDFCAQSPFGYKMSGSCKELCEYGLQAYTEVKTIVTKV 496
 DB 437 VSNLAKAGTVWVNTYNTLHMQPPGKYKESGIGPELGEDALANYTQTKTVSTPI 490

RESULT 10

US-09-344-882-22
 : Sequence 22, Application US/09344882
 : Patent No. US20020162137A1
 : GENERAL INFORMATION:
 : APPLICANT: Nikolau, Basil J
 : APPLICANT: Warteale, Eve S
 : APPLICANT: Oliver, David J
 : APPLICANT: Behal, Robert
 : APPLICANT: Schnable, Patrick S
 : APPLICANT: Ke, Jinsan
 : APPLICANT: Johnson, Jerry L
 : APPLICANT: Allred, Carolyn C

```

; APPLICANT: Fatland, Beth
; APPLICANT: Lutziqer, Isabelle
; APPLICANT: Wen, Tsui-Jung
; TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
; TITLE OF INVENTION: Acetyl CoA Levels in Plants
; FILE REFERENCE: 201573
; CURRENT APPLICATION NUMBER: US/09/344,882
; CURRENT FILING DATE: 1999-06-25
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patent In Ver. 2.2
; SEQ ID NO 22
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Arabidopsis Thaliana
US-09-344-882-22

```

```

Query Match      48.8%; Score 1280; DB 9; Length 501;
Best Local Similarity 53.1%; Pred. No. 3,4e+110;
Matches 255; Conservative 74; Mismatches 149; Indels 2; Gaps 2;

QY 16 PEVPCNOIFINNEHDAVSRKTEPTVNPSTGEVICQVAGSKEDVDKAREGRPGAFQUGS 75
Db 15 PEIKFTKLFNGQFIDAASGKIFETIDPRNGEVIAIAGGKEDVDLAVNAARYAFDHG- 74
QY 76 PWRMDASHSGRLNRLADLIERDRTYLAALETLDNGKPYVISYLVLDLMDVLKCLRYVAG 135
Db 74 PWRMTGFERAKLNLKFALENTTEELAKLDADVAGKIFQLGKYADIPATAGHFRYNAG 133
QY 136 WADKYHCKTIPID-GDFEFSYTHREIPVCGGQIIPNFPFLDMQAMKLGHPALATGNVVMKV 194
Db 134 AADKIHGETLKMTRQSLFGYTLKEPIGVVGNIPWNFTSINEATKVAFAAMAAGTMMVKP 193
QY 195 ABOTPTTALYVANLKEAGFPVGVNIVPGFPTAGAAASHEDVDKVAFTGSTEIGRVI 254
Db 194 ABOTLSALFYAHLKSEAGIPGVNLIVTGFGSTAGAAASHEDVDKVAFTGSTEIGRVI 253
QY 255 QVAAAGSNLRKRVTELGKSPNII NSADUMWAVEQAIFALFENAGQCCAGSKTFVQED 314
Db 254 MQAAAASNLKVKVSELGKSPILLIFENDADIKAADLALLGCFYNNKGEICVASSKVFVQEG 313
QY 315 IYDEFVVRVARAKSRVGNPNDSKTEQGPQVDETFQPKKILGYINTGQEGAKILLCCGGI 374
Db 314 IYDVKVEKIVAKAKMTGDPDSTARQGPQVDETFQPKKILGYINTGQEGAKILLCCGGI 373
QY 375 AADRGYFIQPTVFGVQDGMIIAKKEIFGVPMOILKEKTEIEVGSRRANNSTYGLAAAVET 434
Db 374 IGRKGYFIQPTIFALVTELMKLYGDEISGPVNSLMKTEFVVEGGLKANNIKYGLAAHLS 433
QY 435 KULDKANYLSQALQAGTVWVNCYDVFAGSPRGYKMSGSGRELGEYGLQAYTEKTVTV 494
Db 434 QDIDLINTVSRISKAGITWVNCYFGFDLDCPYGYKMSGSGRESGMALDNYLQTKSVVM 493

```

RESULT 11

```

US-10-293-865-22
; Sequence 22, Application US/10293865
; Publication No. US20030106090A1
; GENERAL INFORMATION:
; APPLICANT: Nikolau, Basil J
; APPLICANT: Wurtele, Eve S
; APPLICANT: Oliver, David J
; APPLICANT: Behal, Robert
; APPLICANT: Schnable, Patrick S
; APPLICANT: Ke, Jinshan
; APPLICANT: Johnson, Jerry L
; APPLICANT: Allred, Carolyn C
; APPLICANT: Fatland, Beth
; APPLICANT: Lutziqer, Isabelle
; APPLICANT: Wen, Tsui-Jung
; TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
; TITLE OF INVENTION: Acetyl CoA Levels in Plants

```

```

; FILE REFERENCE: 217113
; CURRENT APPLICATION NUMBER: US/10/293,865
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/744,892
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 60/090,717
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Patent In Ver. 3.1
; SEQ ID NO 22
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Arabidopsis Thaliana
US-10-293-865-22

```

```

Query Match      48.8%; Score 1280; DB 9; Length 501;
Best Local Similarity 53.1%; Pred. No. 3,4e+110;
Matches 255; Conservative 74; Mismatches 149; Indels 2; Gaps 2;

QY 16 PEVPCNOIFINNEHDAVSRKTEPTVNPSTGEVICQVAGSKEDVDKAREGRPGAFQUGS 75
Db 15 PEIKFTKLFNGQFIDAASGKIFETIDPRNGEVIAIAGGKEDVDLAVNAARYAFDHG- 74
QY 76 PWRMDASHSGRLNRLADLIERDRTYLAALETLDNGKPYVISYLVLDLMDVLKCLRYVAG 135
Db 74 PWRMTGFERAKLNLKFALENTTEELAKLDADVAGKIFQLGKYADIPATAGHFRYNAG 134
QY 136 WADKYHCKTIPID-GDFEFSYTHREIPVCGGQIIPNFPFLDMQAMKLGHPALATGNVVMKV 194
Db 134 AADKIHGETLKMTRQSLFGYTLKEPIGVVGNIPWNFTSINEATKVAFAAMAAGTMMVKP 193
QY 195 ABOTPTTALYVANLKEAGFPVGVNIVPGFPTAGAAASHEDVDKVAFTGSTEIGRVI 254
Db 194 ABOTLSALFYAHLKSEAGIPGVNLIVTGFGSTAGAAASHEDVDKVAFTGSTEIGRVI 253
QY 255 QVAAAGSNLRKRVTELGKSPNII NSADUMWAVEQAIFALFENAGQCCAGSKTFVQED 314
Db 254 MQAAAASNLKVKVSELGKSPILLIFENDADIKAADLALLGCFYNNKGEICVASSKVFVQEG 313
QY 315 IYDEFVVRVARAKSRVGNPNDSKTEQGPQVDETFQPKKILGYINTGQEGAKILLCCGGI 374
Db 314 IYDVKVEKIVAKAKMTGDPDSTARQGPQVDETFQPKKILGYINTGQEGAKILLCCGGI 373
QY 375 AADRGYFIQPTVFGVQDGMIIAKKEIFGVPMOILKEKTEIEVGSRRANNSTYGLAAAVET 434
Db 374 IGRKGYFIQPTIFALVTELMKLYGDEISGPVNSLMKTEFVVEGGLKANNIKYGLAAHLS 433
QY 435 KULDKANYLSQALQAGTVWVNCYDVFAGSPRGYKMSGSGRELGEYGLQAYTEKTVTV 494
Db 434 QDIDLINTVSRISKAGITWVNCYFGFDLDCPYGYKMSGSGRESGMALDNYLQTKSVVM 493

```

RESULT 12

```

US-09-847-208-11
; Sequence 11, Application US/09847208
; Publication No. US20030082190A1
; GENERAL INFORMATION:
; APPLICANT: Saxena, Anand
; APPLICANT: Zhang, Ke
; APPLICANT: Zhu, Daocunq
; TITLE OF INVENTION: Fusion Molecules and Treatment of
; TITLE OF INVENTION: IgE-MEDIATED ALLERGIC DISEASES
; FILE REFERENCE: UC67,002A
; CURRENT APPLICATION NUMBER: US/09/847,208
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 177
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 495
; TYPE: PRT
; ORGANISM: Alternaria alternata
US-09-847-208-11

```


[illegible]

Query Match	45 RA	Score	1201	5	DR	9	Length	492
Best Local Similarity	52.4%	Pred.	No.	6.3e-103				
Matches	258	Conservative	68	Mismatches	143	Indels	23	Gaps
28	FWHFAVSKPTFTVPNST	---FEVTCGVAFTHKLVKAFPEKPEKAPALGSGIHWKMUASHSG	86					
1	EWDSASGKTEVVNFA	NKGEVGRVPEATAEDVDAAVRAAKAEAFKSGPWAKVFAASRA	60					
87	PILNPLADLEPDPYTL	AALETLDINGKPYV--ISLVLDLMVLKCLSYVAWACKYHKGK--	143					
61	PILRLKLDLLEPEDE	LAALETLDLCKPLAFAGSGTEVGPADLEIPIYAGWARKLNGEPP	120					
144	---TIPIDGD--FFSY	THPEVGVCGQIIPNFFPLIMQAKLGPALATGNVVYKVAEQTP	199					
121	VIPSLATGDDELNT	REPLGVGVGISPNFPPLLLAKMLAPALAAAGNTIVLKPSEQTP	180					
209	LTALYVANLKEAG---	FPDGVNIVGEGPTAGAAASHEDVDKVAFTCSIEICRVIQV	256					
181	LTALLAFLEIFFAANF	PKVWNVGSPGAPWGLLSHFLKRTKFTFESIEVSEKLIME	240					
257	ANGSSNLKRVTLGLG	KSPNINSDADIMWAVEGAHFALFFNQGQGVAGSFTFVQEDII	316					
241	AAAANKLKVTLGLG	KSPVIVFDADLDKAVSRIVFGAFGNAGQVCIAPSRLLVHESIV	400					
317	DEPVVRVARAKS--	RVVGNPDPSKTF--QSPVDVDTQPKKILGIYNTKPGFAKILCGGCI	374					
301	PEFVEKLEKWKVKL	KLKLKPLGTSQINILYKLSNQQEIVGLVSIPEKKEEFAKVIQSSPEP	400					

QY	375	AADR-----GYFVGTFVFGWQGGMTIAKRFIPGVNQLKFKFIIFFVWVGRANNSTVYGLA	429
	:	:	
Db	361	DESKYVGGYVQVPTFTFTVTPMKIMKRFIEGVPLEIKFKULLEALANLVEYGLA	429
	:	:	
QY	430	AAVFTKD-LDKANYLSQALQAGTVWN--CYDFVGAQSPFGYKYM-SGSGREL-GEYGLQ	484
	:	:	
Db	421	AVYFTKDLARAFVAKALEAGVWVNDVCVHAAPQLPFGYKQSSGIGREHSGMYGLE	480
	:	:	
QY	485	AYTEVKTVTVKV	496
	:	:	
Db	481	EYTEIKTVTIRL	492
	:	:	

RESULT 15

```

US-10-175-696-21
: Sequence 21, Application US/10175696
: Publication No. US20030092658A1
: GENERAL INFORMATION:
: APPLICANT: Glucksmann, Maria Alexandra
: APPLICANT: Meyers, Rachel
: APPLICANT: Rudolph-Owen, Laura A.
: TITLE OF INVENTION: NOVEL HUMAN ENZYME FAMILY MEMBERS AND USES THEREOF
: FILE REFERENCE: 10448-193001
: CURRENT APPLICATION NUMBER: US/10/175,696
: PRIOR FILING DATE: 2002-05-20
: PRIOR APPLICATION NUMBER: 10/067,668
: PRIOR FILING DATE: 2002-04-04
: PRIOR APPLICATION NUMBER: 60/266,140
: PRIOR FILING DATE: 2001-02-02
: PRIOR APPLICATION NUMBER: 09/843,401
: PRIOR FILING DATE: 2001-03-30
: PRIOR APPLICATION NUMBER: PCT/US01/10720
: PRIOR FILING DATE: 2001-04-02
: PRIOR APPLICATION NUMBER: 60/193,920
: PRIOR FILING DATE: 2000-03-31
: PRIOR APPLICATION NUMBER: 09/862,658
: PRIOR FILING DATE: 2001-05-21
: PRIOR APPLICATION NUMBER: PCT/US01/16380
: PRIOR FILING DATE: 2001-05-21
: PRIOR APPLICATION NUMBER: 60/205,675
: PRIOR FILING DATE: 2000-05-19
: PRIOR APPLICATION NUMBER: 09/882,837
: PRIOR FILING DATE: 2001-06-15
: PRIOR APPLICATION NUMBER: PCT/US01/19319
: PRIOR FILING DATE: 2001-06-15
: PRIOR APPLICATION NUMBER: 60/211,727
: PRIOR FILING DATE: 2000-06-15
: NUMBER OF SEQ ID NOS: 31
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 21
: LENGTH: 493
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: consensus sequence
US-10-175-696-21

```

[illegible]

Search completed: June 24, 2003, 10:37:05
Job time : 26.1692 secs

Genome version 5.1.6
Copyright (c) 1994 - 2002 Compugen Ltd

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:11:00 ; Search time 141.017 seconds
(without alignments)
1780.867 Million cell updates/sec

Title: US-09-830-751-4
Perfect score: 2625
Sequence: 1 MSAAATGAVPAPNQPEVFCYGLQAYTEKIVIVKVPQKN 500

Scoring table: RUSUM62
Gapref in 0, Gapext 0.5

Searched: 4569144 seqs, 64473110 residues
Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Pending_Patent's_AA_Main.*
- 1: /cgn2_6/ptodata/1/paa/US06_COMB.pcp.*
 - 2: /cgn2_6/ptodata/1/paa/US06_COMB.pcp.*
 - 3: /cgn2_6/ptodata/1/paa/US07_COMB.pcp.*
 - 4: /cgn2_6/ptodata/1/paa/US08_COMB.pcp.*
 - 5: /cgn2_6/ptodata/1/paa/US081_COMB.pcp.*
 - 6: /cgn2_6/ptodata/1/paa/US082_COMB.pcp.*
 - 7: /cgn2_6/ptodata/1/paa/US083_COMB.pcp.*
 - 8: /cgn2_6/ptodata/1/paa/US084_COMB.pcp.*
 - 9: /cgn2_6/ptodata/1/paa/US085_COMB.pcp.*
 - 10: /cgn2_6/ptodata/1/paa/US086_COMB.pcp.*
 - 11: /cgn2_6/ptodata/1/paa/US087_COMB.pcp.*
 - 12: /cgn2_6/ptodata/1/paa/US088_COMB.pcp.*
 - 13: /cgn2_6/ptodata/1/paa/US089_COMB.pcp.*
 - 14: /cgn2_6/ptodata/1/paa/US090_COMB.pcp.*
 - 15: /cgn2_6/ptodata/1/paa/US091_COMB.pcp.*
 - 16: /cgn2_6/ptodata/1/paa/US092_COMB.pcp.*
 - 17: /cgn2_6/ptodata/1/paa/US093_COMB.pcp.*
 - 18: /cgn2_6/ptodata/1/paa/US094_COMB.pcp.*
 - 19: /cgn2_6/ptodata/1/paa/US095_COMB.pcp.*
 - 20: /cgn2_6/ptodata/1/paa/US096_COMB.pcp.*
 - 21: /cgn2_6/ptodata/1/paa/US097_COMB.pcp.*
 - 22: /cgn2_6/ptodata/1/paa/US098_COMB.pcp.*
 - 23: /cgn2_6/ptodata/1/paa/US099_COMB.pcp.*
 - 24: /cgn2_6/ptodata/1/paa/US100_COMB.pcp.*
 - 25: /cgn2_6/ptodata/1/paa/US101_COMB.pcp.*
 - 26: /cgn2_6/ptodata/1/paa/US102_COMB.pcp.*
 - 27: /cgn2_6/ptodata/1/paa/US60_COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2624	99.9	516	21	US-09-791-537-77606
2	2574	98.1	516	7	US-08-366-2108-4
3	2574	98.1	515	21	US-09-791-537-75717
4	2572	98.0	517	19	US-09-538-092-851
5	2572	98.0	517	21	US-09-791-537-74079
6	2572	98.0	517	27	US-60-389-987-1242

7	2572	98.0	517	27	US-60-412-418-1242
8	2570	97.9	516	7	US-08-366-2108-4
9	2569	97.9	517	21	US-09-791-537-71773
10	2569	97.9	517	27	US-60-389-987-56
11	2569	97.9	517	27	US-60-412-418-56
12	2544	95.4	494	21	US-09-791-537-91557
13	2503	95.4	500	21	US-09-791-537-64443
14	2500	95.2	516	21	US-09-791-537-74080
15	2480	94.5	519	21	US-09-791-537-130904
16	2447.5	93.2	519	21	US-09-763-443-1247
17	2447.5	93.2	519	21	US-09-760-475-3423
18	2447.5	93.2	519	26	US-10-212-054-1247
19	2436	92.8	494	21	US-09-791-537-19515
20	2436	92.8	520	21	US-09-791-537-74076
21	2434	92.7	500	10	US-08-512-853-13
22	2434	92.7	500	21	US-09-791-537-74077
23	2426	92.4	520	13	US-08-512-853-12
24	2413	91.9	517	21	US-09-791-537-76513
25	2083	74.4	494	21	US-09-791-537-71778
26	1943	74.0	517	27	US-60-389-987-1227
27	1943	74.0	517	27	US-60-412-418-1227
28	1939	73.9	517	21	US-09-791-537-50322
29	1933	73.6	559	21	US-09-760-443-1406
30	1933	73.6	559	21	US-09-760-475-3429
31	1933	73.6	559	26	US-10-212-054-1406
32	1929	73.5	517	1	PCR-US01-17253-3
33	1929	73.5	517	19	US-09-538-092-1047
34	1929	73.5	517	21	US-09-791-537-77611
35	1866	71.1	520	20	US-09-614-150-7212
36	1866	71.1	520	21	US-09-791-537-129281
37	1866	71.1	520	27	US-60-191-681-5649
38	1866	71.1	520	27	US-60-191-681-5649
39	1857	70.7	538	27	US-60-167-217-7335
40	1857	70.7	538	27	US-60-173-464-5881
41	1802	68.6	509	21	US-09-791-537-73639
42	1795	68.5	502	21	US-09-791-537-66277
43	1795	68.4	494	21	US-09-791-537-115118
44	1794	68.3	494	6	US-08-512-853-12
45	1794	68.3	494	21	US-09-791-537-126493

ALIGNMENTS

RESULT 1
US-09-791-537-77606
Sequence 77606, Application US/09791537
GENERAL INFORMATION:
APPLICANT: Biocomix, Inc.
AFFILIANT: Debe, Derek
INVENTOR: Danzel, Joseph
TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
TITLE OF INVENTION: METHODS OF USE THEREOF
FILE REFERENCE: 261/210
CURRENT APPLICATION NUMBER: US/09/791,537
CURRENT FILING DATE: 2001 02 22
NUMBER OF SEQ ID NOS: 154055
SOFTWARE: PatentIn version 3.0
SEQ ID NO 77606
LENGTH: 516
TYPE: PRT
ORGANISM: Homo sapiens
US-09-791-537-77606

Query Match	99.9%	Score	2622	DB	21	Length	516
Best Local Similarity	99.8%	Pred. No.	5e-262				
Matches	494	Conservative	1	Mismatches	0	Indels	0
Q1	1	MSAAATGAVPAPNQPEVFC	INNEWHDAVSKITPTVNFSEVCCVAALGPKEDV	60			
	YGLQAYTEKIVIVKVPQKNYGLQAYTEKIVIVKVPQKN	500			
DB	16	1	MSAAATGAVPAPNQPEVFC	INNEWHDAVSKITPTVNFSEVCCVAALGPKEDV	75		
		YGLQAYTEKIVIVKVPQKNYGLQAYTEKIVIVKVPQKN	500		
Q2	41	US-09-791-537-74079	US-09-791-537-74079	1242			

Db 76 DKAREGRGAFQGSWRMDASHSRLLNRLADLERDTYLAALETLDNCKPVVSYL 135
QY 121 VLDMLVKLYYAGWADKYHGTIPIDGDFSYTRHEPVGCGQIIPWNEPFLMQAWKL 180
Db 136 VLDMLVKLYYAGWADKYHGTIPIDGDFSYTRHEPVGCGQIIPWNEPFLMQAWKL 195
QY 181 GPALATGNVVMKVAQTPLTALYVANLKEAGFPVGVNIVPGFPTAGAAIASHEDVD 240
Db 196 GPALATGNVVMKVAQTPLTALYVANLKEAGFPVGVNIVPGFPTAGAAIASHEDVD 255
QY 241 KVAFTGSTEIGRVQVAAAGSSNLKRVTTLELGKSPNIMSUDMDWAVEQAHFALFPNQG 300
Db 256 KVAFTGSTEIGRVQVAAAGSSNLKRVTTLELGKSPNIMSUDMDWAVEQAHFALFPNQG 315
QY 301 OCCASKRTFVQEGIDYEFVVRKSVARAKSKVGNFUSKTEGSPVDETQPKKILGYINT 360
Db 316 OCCASKRTFVQEGIDYEFVVRKSVARAKSKVGNFUSKTEGSPVDETQPKKILGYINT 375
QY 361 GKQEGAKLLGGGTAADRGYFQPTVFGDVQDGMTIAKEEIPGVNQIILKEKTIPEVVGR 420
Db 376 GKQEGAKLLGGGTAADRGYFQPTVFGDVQDGMTIAKEEIPGVNQIILKEKTIPEVVGR 435
QY 421 ANNSTYGLAAAVFTKDLKANYLSQALQGTVMVNCYDVFAGQSPGGYKMSGSGRFLGE 480
Db 436 ANNSTYGLAAAVFTKDLKANYLSQALQGTVMVNCYDVFAGQSPGGYKMSGSGRFLGE 495
QY 481 YGLQAYTEVKTIVTKVPQKN 500
Db 496 YGLQAYTEVKTIVTKVPQKN 515

RESULT 2

US-08-466-210B-2
; Sequence 2, Application US/08366210B
; GENERAL INFORMATION:
; APPLICANT: Lauerman et al.
; TITLE OF INVENTION: Expression of a Mutant Form of Aldehyde
; TITLE OF INVENTION: dehydrogenase as an Alcohol Aversive Agent
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/366,210B
; FILING DATE: December 29, 1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane E. Remillard
; REGISTRATION NUMBER: 38,872
; REFERENCE/DOCKET NUMBER: TTI-128
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 515 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-466-210B 2

Query Match 98.1%; Score 2574, DB 7, Length 515,
Best Local Similarity 99.8%; Pred. No. 4.8e-257;

Matches 489; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 11 APNQOPEVFCNQIFINNEHDAVSRKTTPTVNPSTGEVTCQVAVGKREIWKAREBRGA 70
Db 25 APNQOPEVFCNQIFINNEHDAVSRKTTPTVNPSTGEVTCQVAVGKREIWKAREBRGA 84
QY 71 FOLGSPWRMDASHSRLLNRLADLERDTYLAALETLDNCKPVVSYLVLDMLVKL 140
Db 85 FOLGSPWRMDASHSRLLNRLADLERDTYLAALETLDNCKPVVSYLVLDMLVKL 144
QY 131 RYVAGWADKYHGTIPIDGDFSYTRHEPVGCGQIIPWNEPFLMQAWKLGFALATGNV 190
Db 145 RYVAGWADKYHGTIPIDGDFSYTRHEPVGCGQIIPWNEPFLMQAWKLGFALATGNV 204
QY 191 VMKVAQTPLTALYVANLKEAGFPVGVNIVPGFPTAGAAIASHEDVDKVAFTGSTET 250
Db 205 VMKVAQTPLTALYVANLKEAGFPVGVNIVPGFPTAGAAIASHEDVDKVAFTGSTET 264
QY 251 GRVIOVAAAGSSNLKRVTTLELGKSPNIMSUDMDWAVEQAHFALFPNQGTCVAGSKTF 310
Db 265 GRVIOVAAAGSSNLKRVTTLELGKSPNIMSUDMDWAVEQAHFALFPNQGTCVAGSKTF 324
QY 311 VQPIVYFVFWVSVAARAKSKVGNFUSKTEGSPVDETQPKKILGYINTKQEGAKLL 370
Db 325 VQPIVYFVFWVSVAARAKSKVGNFUSKTEGSPVDETQPKKILGYINTKQEGAKLL 384
QY 371 GGGTAADRGYFQPTVFGDVQDGMTIAKEEIPGVNQIILKEKTIPEVVGRANNSTYGLAA 430
Db 385 GGGTAADRGYFQPTVFGDVQDGMTIAKEEIPGVNQIILKEKTIPEVVGRANNSTYGLAA 444
QY 431 AVFTKDLKANYLSQALQGTVMVNCYDVFAGQSPGGYKMSGSGRFLGEYGLAYTEVK 490
Db 445 AVFTKDLKANYLSQALQGTVMVNCYDVFAGQSPGGYKMSGSGRFLGEYGLAYTEVK 504
QY 491 TVTKVPQKN 500
Db 505 TVTKVPQKN 514

RESULT 3

US-09-791-537-57517
; Sequence 57517, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Biomimix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Panzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001 02 22
; NUMBER OF SEQ ID NOS: 154055
; SOFTWARE: Patent In version 4.0
; SEQ ID NO 57517
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-791-537-57517

Query Match 98.1%; Score 2574; DB 21; Length 515;
Best Local Similarity 99.8%; Pred. No. 4.8e-257;

Matches 489; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 11 APNQOPEVFCNQIFINNEHDAVSRKTTPTVNPSTGEVTCQVAVGKREIWKAREBRGA 70
Db 25 APNQOPEVFCNQIFINNEHDAVSRKTTPTVNPSTGEVTCQVAVGKREIWKAREBRGA 84
QY 71 FOLGSPWRMDASHSRLLNRLADLERDTYLAALETLDNCKPVVSYLVLDMLVKL 140
Db 85 FOLGSPWRMDASHSRLLNRLADLERDTYLAALETLDNCKPVVSYLVLDMLVKL 144
QY 131 RYVAGWADKYHGTIPIDGDFSYTRHEPVGCGQIIPWNEPFLMQAWKLGFALATGNV 190
Db 145 RYVAGWADKYHGTIPIDGDFSYTRHEPVGCGQIIPWNEPFLMQAWKLGFALATGNV 204

```

Db 145 PYYAGWADKYHKTIPITIGDFESSYIRHEPVGVVQSLIPNNPFLMLAWKLGALALGNVV 204
QY 191 VMKVAQTPLTALYVANLKEAGFPFGVNVIRPSPATAGAAIASHEWIKVAFQDSIRP 250
Db 205 VMKVAQTPLTALYVANLKEAGFPFGVNVIRPSPATAGAAIASHEWIKVAFQDSIRP 264
QY 251 GRVTVAAAGSSNLKRPVTLLEGGKSPNII MSDALIMWAVELAHFALFENQSGGASSTP 410
Db 265 GRVTVAAAGSSNLKRPVTLLEGGKSPNII MSDALIMWAVELAHFALFENQSGGASSTP 424
QY 311 VQEDIDYEFVVPVSAPAKSPVGNPFTSKTEQSGQVDETFQFKILGYINTGQESAKLLA 370
Db 325 VQEDIDYEFVVPVSAPAKSPVGNPFTSKTEQSGQVDETFQFKILGYINTGQESAKLLA 384
QY 371 GGGIAADRGYFIQPTVFGVQDGMTTAKBEIFSPVWQILKFKTIEEVVSGANNISYGLAA 430
Db 385 GGGIAADRGYFIQPTVFGVQDGMTTAKBEIFSPVWQILKFKTIEEVVSGANNISYGLAA 444
QY 431 AVFTKDLDKANYLSOALQAGTVWVNCYDVFAGSPGSGYKMSGSPGELGEYGLQAYTEVK 490
Db 445 AVFTKDLDKANYLSOALQAGTVWVNCYDVFAGSPGSGYKMSGSPGELGEYGLQAYTEVK 504
QY 491 TTVTKVPQKN 500
Db 505 TTVTKVPQKN 514

```

RESULT 4

```

US-09-538-092-851
: Sequence 851, Application US/09538092
: GENERAL INFORMATION:
: APPLICANT: Mansfield, Traci A.
: TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
: FILE REFERENCE: 15966-542
: CURRENT APPLICATION NUMBER: US/09/538,092
: PRIOR FILING DATE: 2000-04-29
: PRIOR APPLICATION NUMBER: 60/137,352
: PRIOR FILING DATE: 1999-04-01
: PRIOR APPLICATION NUMBER: 60/178,965
: PRIOR FILING DATE: 2000-02-01
: NUMBER OF SEQ ID NOS: 138?
: SEQ ID NO 851
: SOFTWARE: GigaFats2/Formatter Version 0.9
: LENGTH: 517
: TYPE: PRT
: ORGANISM: Homo sapiens
: NAME/KEY: misc_feature
: LOCATION: (0) (0)
: OTHER INFORMATION: Polypeptide Accession Number P05091
US-09-538-092-851

```

```

Query Match 98.0%, Score 2572, DB 19, Length 517,
Best Local Similarity 98.2%, Pred. No. 7.8e-257;
Matches 491, Conservative 2, Mismatches 7, Indels 0, Gaps 0;

QY 1 MSAAATQAVPAPNQPEVFNQIFINNEHDAVSRKTEPTVNFSTGEVLCQVABGJKEVD 60
Db 17 LSAAATQAVPAPNQPEVFNQIFINNEHDAVSRKTEPTVNFSTGEVLCQVABGJKEVD 76
QY 61 DKAREGRGAFQLGSPWPPMDASHSPFLNPLADLIEPDPTYLAAIFTLNPKPVVISYL 120
Db 77 DKAVKAARAAAFQLGSPWPPMDASHSPFLNPLADLIEPDPTYLAAIFTLNPKPVVISYL 136
QY 121 VLLDMVLKCLRYTAGWADKYHGKTIPLDGFSSYIRHEPVGVVQSLIPNNPFLMLAWKLG 180
Db 137 VLLDMVLKCLRYTAGWADKYHGKTIPLDGFSSYIRHEPVGVVQSLIPNNPFLMLAWKLG 196
QY 181 GPALATGNVVMKVAEQTPLTALYVANLKEAGFPFGVNVIRPSPATAGAAIASHEWID 240
Db 197 GPALATGNVVMKVAEQTPLTALYVANLKEAGFPFGVNVIRPSPATAGAAIASHEWID 256
QY 241 KVAFTGSTEIRPVIQVAAAGSSNLKRPVTLLEGGKSPNII MSDALIMWAVELAHFALFENQ 300
Db 257 KVAFTGSTEIRPVIQVAAAGSSNLKRPVTLLEGGKSPNII MSDALIMWAVELAHFALFENQ 316
QY 361 GKKPEAKLIDGGSSLAAGSYFETQFVFGWAGSMTLAKBEIFSPVWQILKFKTIEEVVSR 420
Db 377 GKKPEAKLIDGGSSLAAGSYFETQFVFGWAGSMTLAKBEIFSPVWQILKFKTIEEVVSR 436

```

```

QY 241 KVAFTGSTEIRPVIQVAAAGSSNLKRPVTLLEGGKSPNII MSDALIMWAVELAHFALFENQ 300
Db 257 KVAFTGSTEIRPVIQVAAAGSSNLKRPVTLLEGGKSPNII MSDALIMWAVELAHFALFENQ 316
QY 361 GKKPEAKLIDGGSSLAAGSYFETQFVFGWAGSMTLAKBEIFSPVWQILKFKTIEEVVSR 420
Db 377 GKKPEAKLIDGGSSLAAGSYFETQFVFGWAGSMTLAKBEIFSPVWQILKFKTIEEVVSR 436
QY 481 ANNSTYLAAAVETKLDYAKYLSALDACLWVWNYGVFGASLSPHSGYFYMSSGSGRELGE 480
Db 497 ANNSTYLAAAVETKLDYAKYLSALDACLWVWNYGVFGASLSPHSGYFYMSSGSGRELGE 496
QY 491 YGLQAYTEVKTIVTKVPQKN 500
Db 497 YGLQAYTEVKTIVTKVPQKN 514

RESULT 5
US-09-791-537-74079
: Sequence 74079, Application US/09791537
: GENERAL INFORMATION:
: APPLICANT: Biocomix, Inc.
: APPLICANT: Debe, Derek
: APPLICANT: Danzer, Joseph
: TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
: TITLE OF INVENTION: METHODS OF USE THEREOF
: FILE REFERENCE: 261/210
: CURRENT APPLICATION NUMBER: US/09/791,537
: CURRENT FILING DATE: 2001-02-22
: NUMBER OF SEQ ID NOS: 153055
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 74079
: LENGTH: 517
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-791-537-74079

Query Match 98.0%, Score 2572, DB 21, Length 517;
Best Local Similarity 98.2%, Pred. No. 7.8e-257;
Matches 491, Conservative 2, Mismatches 7, Indels 0, Gaps 0;

QY 1 MSAAATQAVPAPNQPEVFNQIFINNEHDAVSRKTEPTVNFSTGEVLCQVABGJKEVD 60
Db 17 LSAAATQAVPAPNQPEVFNQIFINNEHDAVSRKTEPTVNFSTGEVLCQVABGJKEVD 76
QY 61 DKAREGRGAFQLGSPWPPMDASHSPFLNPLADLIEPDPTYLAAIFTLNPKPVVISYL 120
Db 77 DKAVKAARAAAFQLGSPWPPMDASHSPFLNPLADLIEPDPTYLAAIFTLNPKPVVISYL 136
QY 121 VLLDMVLKCLRYTAGWADKYHGKTIPLDGFSSYIRHEPVGVVQSLIPNNPFLMLAWKLG 180
Db 137 VLLDMVLKCLRYTAGWADKYHGKTIPLDGFSSYIRHEPVGVVQSLIPNNPFLMLAWKLG 196
QY 181 GPALATGNVVMKVAEQTPLTALYVANLKEAGFPFGVNVIRPSPATAGAAIASHEWID 240
Db 197 GPALATGNVVMKVAEQTPLTALYVANLKEAGFPFGVNVIRPSPATAGAAIASHEWID 256
QY 241 KVAFTGSTEIRPVIQVAAAGSSNLKRPVTLLEGGKSPNII MSDALIMWAVELAHFALFENQ 300
Db 257 KVAFTGSTEIRPVIQVAAAGSSNLKRPVTLLEGGKSPNII MSDALIMWAVELAHFALFENQ 316
QY 361 GKKPEAKLIDGGSSLAAGSYFETQFVFGWAGSMTLAKBEIFSPVWQILKFKTIEEVVSR 420
Db 377 GKKPEAKLIDGGSSLAAGSYFETQFVFGWAGSMTLAKBEIFSPVWQILKFKTIEEVVSR 436

```

QY 421 ANNSTYGLAAAVTKDLKANKYLSQALQAGTVWVNCYDVFAGASPPGCGYKMGSGSRELGE 480
 Db 437 ANNSTYGLAAAVTKDLKANKYLSQALQAGTVWVNCYDVFAGASPPGCGYKMGSGSRELGE 495
 QY 481 YGLQAYTEVKTIVTKVPQKN 500
 Db 497 YGLQAYTEVKTIVTKVPQKN 516

RESULT 6

US-60-389-987-1242
 ; Sequence 1242, Application US/6038987
 ; GENERAL INFORMATION:
 ; APPLICANT: Ghosh, Soumitra S.
 ; APPLICANT: Fahy, Eoin D.
 ; APPLICANT: Zhang, Bing
 ; APPLICANT: Gibson, Bradford W.
 ; APPLICANT: Taylor, Steven W.
 ; APPLICANT: Glenn, Gary M.
 ; APPLICANT: Warnock, Dale E.
 ; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
 ; FILE REFERENCE: 660088.465P2
 ; CURRENT APPLICATION NUMBER: US/60/389,987
 ; CURRENT FILING DATE: 2002-06-17
 ; NUMBER OF SEQ ID NOS: 3025
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1242
 ; LENGTH: 517
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-60-389-987-1242

Query Match 98.0%; Score 2572; DB 27; Length 517;
 Best Local Similarity 98.2%; Pred. No. 7.8e-257;
 Matches 491; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 MSAATQAVPAPNQOPEVFCNQIFINNEWHDAVSRTFTVNPSTGEVICOVAEGDKEDV 60
 Db 17 LSAAATQAVPAPNQOPEVFCNQIFINNEWHDAVSRTFTVNPSTGEVICOVAEGDKEDV 76
 QY 61 DKAREKPGAFQUGSPWRPMIAASHSKLLNRLADLIERDTYLAALLETLDNGKPYVSYL 120
 Db 77 DKAVKAARAAAFQUGSPWRPMIAASHGRLLNRLADLIERDTYLAALLETLDNGKPYVSYL 136
 QY 121 VLDLMDVLKCLRYAYAGWADKYHGKTPIDQDFESYTHPEPVVCGQIIPWNEPILMQAWKL 180
 Db 137 VLDLMDVLKCLRYAYAGWADKYHGKTPIDQDFESYTHPEPVVCGQIIPWNEPILMQAWKL 196
 QY 181 GPALATGNVVMKVAEQTPITAIYVANLIKEAGFPFGVNVIVPGPTAGAAIASHEIDV 240
 Db 197 GPALATGNVVMKVAEQTPITAIYVANLIKEAGFPFGVNVIVPGPTAGAAIASHEIDV 256
 QY 241 KVAFTGSTELGRVIGVAAAGSSNLKPVTLFTGKSPNIIMSDAMIMWAVEQAHEALFTNQ 400
 Db 257 KVAFTGSTELGRVIGVAAAGSSNLKPVTLFTGKSPNIIMSDAMIMWAVEQAHEALFTNQ 416
 QY 301 QCCAGSRTFVQEDIVDEPVERSVARAKSRVGNPFDSKTFQGPGVDTGPKKIIIGYINT 360
 Db 317 QCCAGSRTFVQEDIVDEPVERSVARAKSRVGNPFDSKTFQGPGVDTGPKKIIIGYINT 376
 QY 361 KQEGAKLLCGGTAADRGYFTQPTVFGVQDQGMTIAKEETPGVMQILKEKTEIEVVGR 420
 Db 376 KQEGAKLLCGGTAADRGYFTQPTVFGVQDQGMTIAKEETPGVMQILKEKTEIEVVGR 436
 QY 421 ANNSTYGLAAAVTKDLKANKYLSQALQAGTVWVNCYDVFAGASPPGCGYKMGSGSRELGE 480
 Db 437 ANNSTYGLAAAVTKDLKANKYLSQALQAGTVWVNCYDVFAGASPPGCGYKMGSGSRELGE 495
 QY 481 YGLQAYTEVKTIVTKVPQKN 500
 Db 497 YGLQAYTEVKTIVTKVPQKN 516

RESULT 7

US-60-412-418-1242
 ; Sequence 1242, Application US/60412418
 ; GENERAL INFORMATION:
 ; APPLICANT: Ghosh, Soumitra S.
 ; APPLICANT: Fahy, Eoin D.
 ; APPLICANT: Zhang, Bing
 ; APPLICANT: Gibson, Bradford W.
 ; APPLICANT: Taylor, Steven W.
 ; APPLICANT: Glenn, Gary M.
 ; APPLICANT: Warnock, Dale E.
 ; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
 ; FILE REFERENCE: 660088.465P4
 ; CURRENT APPLICATION NUMBER: US/60/412,418
 ; CURRENT FILING DATE: 2002-09-20
 ; NUMBER OF SEQ ID NOS: 3025
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1242
 ; LENGTH: 517
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-60-412-418-1242

Query Match 98.0%; Score 2572; DB 27; Length 517;
 Best Local Similarity 98.2%; Pred. No. 7.8e-257;
 Matches 491; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 MSAATQAVPAPNQOPEVFCNQIFINNEWHDAVSRTFTVNPSTGEVICOVAEGDKEDV 60
 Db 17 LSAAATQAVPAPNQOPEVFCNQIFINNEWHDAVSRTFTVNPSTGEVICOVAEGDKEDV 76
 QY 61 DKAREKPGAFQUGSPWRPMIAASHSKLLNRLADLIERDTYLAALLETLDNGKPYVSYL 120
 Db 77 DKAVKAARAAAFQUGSPWRPMIAASHGRLLNRLADLIERDTYLAALLETLDNGKPYVSYL 136
 QY 121 VLDLMDVLKCLRYAYAGWADKYHGKTPIDQDFESYTHPEPVVCGQIIPWNEPILMQAWKL 180
 Db 137 VLDLMDVLKCLRYAYAGWADKYHGKTPIDQDFESYTHPEPVVCGQIIPWNEPILMQAWKL 196
 QY 181 GPALATGNVVMKVAEQTPITAIYVANLIKEAGFPFGVNVIVPGPTAGAAIASHEIDV 240
 Db 197 GPALATGNVVMKVAEQTPITAIYVANLIKEAGFPFGVNVIVPGPTAGAAIASHEIDV 256
 QY 241 KVAFTGSTELGRVIGVAAAGSSNLKPVTLFTGKSPNIIMSDAMIMWAVEQAHEALFTNQ 400
 Db 257 KVAFTGSTELGRVIGVAAAGSSNLKPVTLFTGKSPNIIMSDAMIMWAVEQAHEALFTNQ 416
 QY 301 QCCAGSRTFVQEDIVDEPVERSVARAKSRVGNPFDSKTFQGPGVDTGPKKIIIGYINT 360
 Db 317 QCCAGSRTFVQEDIVDEPVERSVARAKSRVGNPFDSKTFQGPGVDTGPKKIIIGYINT 376
 QY 361 KQEGAKLLCGGTAADRGYFTQPTVFGVQDQGMTIAKEETPGVMQILKEKTEIEVVGR 420
 Db 376 KQEGAKLLCGGTAADRGYFTQPTVFGVQDQGMTIAKEETPGVMQILKEKTEIEVVGR 436
 QY 421 ANNSTYGLAAAVTKDLKANKYLSQALQAGTVWVNCYDVFAGASPPGCGYKMGSGSRELGE 480
 Db 437 ANNSTYGLAAAVTKDLKANKYLSQALQAGTVWVNCYDVFAGASPPGCGYKMGSGSRELGE 495
 QY 481 YGLQAYTEVKTIVTKVPQKN 500
 Db 497 YGLQAYTEVKTIVTKVPQKN 516

RESULT 8

US-08-466-210B-4
 ; Sequence 4, Application US/08466210B
 ; GENERAL INFORMATION:
 ; APPLICANT: Lauterman et al.
 ; TITLE OF INVENTION: Expression of a Mutant Form of Alcoholdehyde
 ; TITLE OF INVENTION: Dehydrogenase as an Alcohol Acetivase Agent

: NUMBER OF SEQUENCES: 4
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: LAHIVE & CUCKFIELD
 : STREET: 60 State Street, suite 510
 : CITY: Boston
 : STATE: Massachusetts
 : COUNTRY: USA
 : ZIP: 02109 1875
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: ASCII Text
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/366,210B
 : FILING DATE: December 29, 1994
 : CLASSIFICATION: 514
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Jane E. Remillard
 : REGISTRATION NUMBER: 38,872
 : REFERENCE/DOCKET NUMBER: TTI-128
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (617)227-7400
 : TELEFAX: (617)227-5941
 : INFORMATION FOR SEQ ID NO: 4:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 515 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : US-08-366-210R-4

Query Match 97.9%, Score 2570, DB 7, Length 515,
 Best Local Similarity 99.6%, Pred. No. 1.3e-256;
 Matches 488; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 11 APNQPEVFCNQIFINNEHDAVSPKTFPTVNPSTGEVTCQVAEGRKENVKAPGPPGA 70
 DB 25 APNQPEVFCNQIFINNEHDAVSPKTFPTVNPSTGEVTCQVAEGRKENVKAPGPPGA 84
 QY 71 FQAGSPWPMIAASHGPIINPLADIPTDPTTYLALETLENGRPYVSYLVLDWVKLC 130
 DB 85 FQAGSPWPMIAASHGPIINPLADIPTDPTTYLALETLENGRPYVSYLVLDWVKLC 144
 QY 131 RYAGWAHKKYHKTPTPTGTFEFSYTPHEPWSVVGQITPWNFPLLMKAWKLGPAIATSNVV 190
 DB 145 RYAGWAHKKYHKTPTPTGTFEFSYTPHEPWSVVGQITPWNFPLLMKAWKLGPAIATSNVV 204
 QY 191 VMKVAFTPTPTALYVANIKEAGPPGVVNIIVPGFPTAGAAATASHHEVVKVAFSTEL 250
 DB 205 VMKVAEQTPALYVANIKEAGPPGVVNIIVPGFPTAGAAATASHHEVVKVAFSTEL 264
 QY 261 GPVTVAAGSSNI:KPVTLFTGSKSFNI:MSDALM:WAVEKAHFALEFNGQCCAGSPTE 310
 DB 265 GPVTVAAGSSNI:KRVTLFTGSKSFNI:MSDALM:WAVEKAHFALEFNGQCCAGSPTE 324
 QY 311 VQEDYVGEFVVSVAFAKSRVWVNPUSKTEQSPQVHELEPKKII:YINTK:KGEAKILC 470
 DB 325 VQEDYVGEFVVSVAFAKSRVWVNPUSKTEQSPQVHELEPKKII:YINTK:KGEAKILC 384
 QY 371 GGGIAADRGYFIQPTVEGDQGGMTIAKEETFGPMQILAKFKTIEEVVGRANNSTYGLAA 430
 DB 385 GGGIAADRGYFIQPTVEGDQGGMTIAKEETFGPMQILAKFKTIEEVVGRANNSTYGLAA 444
 QY 431 AVFTKDLKANKYLSQALQAGTVMVNCYDVGCAQSPFGGKMSGSGRELGEYGLQAYTEVK 490
 DB 445 AVFTKDLKANKYLSQALQAGTVMVNCYDVGCAQSPFGGKMSGSGRELGEYGLQAYTEVK 504
 QY 491 TVTVKVPQKN 500
 DB 505 TVTVKVPQKN 514

RESULT 9

US-09-791-537-71773
 : Sequence 11/14, Application US/09791537
 : GENERAL INFORMATION:
 : APPLICANT: Bionomix, Inc.
 : APPLICANT: Debe, Derek
 : APPLICANT: Danzer, Joseph
 : TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
 : TITLE OF INVENTION: METHODS OF USE THEREOF
 : FILE REFERENCE: 261/210
 : CURRENT APPLICATION NUMBER: US/09/791,537
 : CURRENT FILING DATE: 2001-02-22
 : NUMBER OF SEQ ID NOS: 153055
 : SOFTWARE: PatentIn version 3.0
 : SEQ ID NO 71773
 : LENGTH: 517
 : TYPE: PRT
 : ORGANISM: Homo sapiens
 : US-09-791-537-71773

Query Match 97.9%, Score 2569, DB 21, Length 517,
 Best Local Similarity 98.9%, Pred. No. 1.6e 256;
 Matches 490, Conservative 3; Mismatches 7; Indels 0; Gaps 0;
 QY 1 MSAAATQAVPAPNQPEVFCNQIFINNEHDAVSPKTFPTVNPSTGEVTCQVAEGRKENV 60
 DB 17 LSAAATQAVPAPNQPEVFCNQIFINNEHDAVSPKTFPTVNPSTGEVTCQVAEGRKENV 76
 QY 61 DKAFEGPHGATGSGSPWPMIAASHGPIINPLADIPTDPTTYLALETLENGRPYVSYLV 120
 DB 77 DKAFVFAAPFQAGSPWPMIAASHGPIINPLADIPTDPTTYLALETLENGRPYVSYLV 136
 QY 121 VQEDWVKLC:RYAGWAHKKYHKTPTPTGTFEFSYTPHEPWSVVGQITPWNFPLLMKAWK 180
 DB 137 VQEDWVKLC:RYAGWAHKKYHKTPTPTGTFEFSYTPHEPWSVVGQITPWNFPLLMKAWK 196
 QY 181 GPALATGNNVVMKVAEFTPTALYVANIKEAGPPGVVNIIVPGFPTAGAAATASHHEV 240
 DB 197 GPALATGNNVVMKVAEFTPTALYVANIKEAGPPGVVNIIVPGFPTAGAAATASHHEV 256
 QY 241 KVAFTGSTEIGPVTCQVAAGSSNI:KPVTLFTGSKSFNI:MSDALM:WAVEKAHFALEFNG 300
 DB 257 KVAFTGSTEIGPVTCQVAAGSSNI:KPVTLFTGSKSFNI:MSDALM:WAVEKAHFALEFNG 316
 QY 301 QCCAGSPTEFVQEDYVGEFVVSVAFAKSRVWVNPUSKTEQSPQVHELEPKKII:GYINT 360
 DB 317 QCCAGSPTEFVQEDYVGEFVVSVAFAKSRVWVNPUSKTEQSPQVHELEPKKII:GYINT 376
 QY 361 QCCAGSPTEFVQEDYVGEFVVSVAFAKSRVWVNPUSKTEQSPQVHELEPKKII:GYINT 420
 DB 377 QCCAGSPTEFVQEDYVGEFVVSVAFAKSRVWVNPUSKTEQSPQVHELEPKKII:GYINT 436
 QY 431 ANNSYGLAAAVSF:KLLKANKYLSQALQAGTVMVNCYDVGCAQSPFGGKMSGSGRELGE 480
 DB 447 ANNSYGLAAAVSF:KLLKANKYLSQALQAGTVMVNCYDVGCAQSPFGGKMSGSGRELGE 496
 QY 481 YGLQAYTEVKTVTVKVPQKN 500
 DB 497 YGLQAYTEVKTVTVKVPQKN 516

RESULT 10

US-50-389-987-56
 : Sequence 56, Application US/60389987
 : GENERAL INFORMATION:
 : APPLICANT: Ghosh, Soumitra S.
 : APPLICANT: Paly, Eoin D.
 : APPLICANT: Zhang, Bing
 : APPLICANT: Gibson, Bradford W.
 : APPLICANT: Taylor, Steven W.
 : APPLICANT: Glenn, Gary M.
 : APPLICANT: Warnock, Dale E.
 : TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION

; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
 ; FILE REFERENCE: 660088.465P2
 ; CURRENT APPLICATION NUMBER: US/60/389,987
 ; CURRENT FILING DATE: 2002-06-17
 ; NUMBER OF SEQ ID NOS: 3025
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 56
 ; LENGTH: 517
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-60-389-987-56

Query Match 97.9%; Score 2569; DB 27; Length 517;
 Best Local Similarity 98.0%; Pred. No. 1.6e-256;
 Matches 490; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
 QY 1 MSAATQAVPAPNQOPEVFCNQIFINNEWHDAVSRRKTEFTVNPSTGEVICOVAGSKEDV 60
 DB 17 LSAATQAVPAPNQOPEVFCNQIFINNEWHDAVSRRKTEFTVNPSTGEVICOVAGSKEDV 76
 QY 61 DKAREGRPGAFQIGSPWRMDASHGRLLNRLADLIERDPTYLAALETLDNGKPVVISYL 120
 DB 77 DKAVKAARAFAQIGSPWRMDASHGRLLNRLADLIERDPTYLAALETLDNGKPVVISYL 136
 QY 121 VLDLMDVLKCLRYVAGWADKYHGKTIPIDGDFEFSYTRHEPVGVCGQIIPWNEPPLLMAQWKL 180
 DB 137 VLDLMDVLKCLRYVAGWADKYHGKTIPIDGDFEFSYTRHEPVGVCGQIIPWNEPPLLMAQWKL 196
 QY 181 GPALATGNVVMKVAQQTPLTALYVANLIKAGFPFGVNVIVPGFPTAGAAIASHEDVD 240
 DB 197 GPALATGNVVMKVAQQTPLTALYVANLIKAGFPFGVNVIVPGFPTAGAAIASHEDVD 256
 QY 241 KVAFTGSTELGRVIOVAAAGSSNKKVTLLELGKSPNIIIMSDADMDWAVEQAHFALFENOG 300
 DB 257 KVAFTGSTELGRVIOVAAAGSSNKKVTLLELGKSPNIIIMSDADMDWAVEQAHFALFENOG 316
 QY 301 QCCAGSRFTVQELIYDEFEVRSVARAKSRVVGNNPFDSEKTEGPGVDETPKKTILGYINT 360
 DB 317 QCCAGSRFTVQELIYDEFEVRSVARAKSRVVGNNPFDSEKTEGPGVDETPKKTILGYINT 376
 QY 361 GKQSGAKLLCGGSIADRGYFTOPTVFGDVQVQMTIAKEEIPGVPMQILKKTIEEVVGR 420
 DB 377 GKQSGAKLLCGGSIADRGYFTOPTVFGDVQVQMTIAKEEIPGVPMQILKKTIEEVVGR 436
 QY 421 ANNSTYGLAAAVFTKDLKANYLSQALQAGTWWVNCYDVFQAQSPFGGKMGSGSREIGE 480
 DB 437 ANNSTYGLAAAVFTKDLKANYLSQALQAGTWWVNCYDVFQAQSPFGGKMGSGSREIGE 496
 QY 481 YGLQAYTEVKTIVTKVPQKN 500
 DB 497 YGLQAYTEVKTIVTKVPQKN 516

RESULT 11

US-60-412-418-56

; Sequence 56, Application US/60412418
 ; GENERAL INFORMATION:
 ; APPLICANT: Ghosh, Soumitra S.
 ; APPLICANT: Pahi, Eoin D.
 ; APPLICANT: Zhang, Bing
 ; APPLICANT: Gibson, Bradford W.
 ; APPLICANT: Taylor, Steven W.
 ; APPLICANT: Glenn, Gary M.
 ; APPLICANT: Wainock, Dale E.
 ; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
 ; FILE REFERENCE: 660088.465P3
 ; CURRENT APPLICATION NUMBER: US/60/412,418
 ; CURRENT FILING DATE: 2002-09-20
 ; NUMBER OF SEQ ID NOS: 3025
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 56
 ; LENGTH: 517

; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-60-412-418-56

Query Match 97.9%; Score 2569; DB 27; Length 517;
 Best Local Similarity 98.0%; Pred. No. 1.6e-256;
 Matches 490; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 MSAATQAVPAPNQOPEVFCNQIFINNEWHDAVSRRKTEFTVNPSTGEVICOVAGSKEDV 60
 DB 17 LSAATQAVPAPNQOPEVFCNQIFINNEWHDAVSRRKTEFTVNPSTGEVICOVAGSKEDV 76
 QY 61 DKAREGRPGAFQIGSPWRMDASHGRLLNRLADLIERDPTYLAALETLDNGKPVVISYL 120
 DB 77 DKAVKAARAFAQIGSPWRMDASHGRLLNRLADLIERDPTYLAALETLDNGKPVVISYL 136
 QY 121 VLDLMDVLKCLRYVAGWADKYHGKTIPIDGDFEFSYTRHEPVGVCGQIIPWNEPPLLMAQWKL 180
 DB 137 VLDLMDVLKCLRYVAGWADKYHGKTIPIDGDFEFSYTRHEPVGVCGQIIPWNEPPLLMAQWKL 196
 QY 181 GPALATGNVVMKVAQQTPLTALYVANLIKAGFPFGVNVIVPGFPTAGAAIASHEDVD 240
 DB 197 GPALATGNVVMKVAQQTPLTALYVANLIKAGFPFGVNVIVPGFPTAGAAIASHEDVD 256
 QY 241 KVAFTGSTELGRVIOVAAAGSSNKKVTLLELGKSPNIIIMSDADMDWAVEQAHFALFENOG 300
 DB 257 KVAFTGSTELGRVIOVAAAGSSNKKVTLLELGKSPNIIIMSDADMDWAVEQAHFALFENOG 316
 QY 301 QCCAGSRFTVQELIYDEFEVRSVARAKSRVVGNNPFDSEKTEGPGVDETPKKTILGYINT 360
 DB 317 QCCAGSRFTVQELIYDEFEVRSVARAKSRVVGNNPFDSEKTEGPGVDETPKKTILGYINT 376
 QY 361 GKQSGAKLLCGGSIADRGYFTOPTVFGDVQVQMTIAKEEIPGVPMQILKKTIEEVVGR 420
 DB 377 GKQSGAKLLCGGSIADRGYFTOPTVFGDVQVQMTIAKEEIPGVPMQILKKTIEEVVGR 436
 QY 421 ANNSTYGLAAAVFTKDLKANYLSQALQAGTWWVNCYDVFQAQSPFGGKMGSGSREIGE 480
 DB 437 ANNSTYGLAAAVFTKDLKANYLSQALQAGTWWVNCYDVFQAQSPFGGKMGSGSREIGE 496
 QY 481 YGLQAYTEVKTIVTKVPQKN 500
 DB 497 YGLQAYTEVKTIVTKVPQKN 516

RESULT 12

US-09-791-537-69657

Sequence 69657, Application US/09791537

; GENERAL INFORMATION:
 ; APPLICANT: Biocomix, Inc.
 ; APPLICANT: Debe, Derek
 ; APPLICANT: Banzer, Joseph
 ; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
 ; FILE REFERENCE: 251/210
 ; CURRENT APPLICATION NUMBER: US/09/791,537
 ; CURRENT FILING DATE: 2001-02-22
 ; NUMBER OF SEQ ID NOS: 153055
 ; SOFTWARE: patentIn version 3.0
 ; SEQ ID NO 69657
 ; LENGTH: 494
 ; TYPE: PRT
 ; ORGANISM: pdb 1CW3A
 ; US-09-791-537-69657

Query Match 95.5%; Score 2544; DB 21; Length 494;
 Best Local Similarity 98.4%; Pred. No. 5.8e-254;
 Matches 485; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 8 AVPAIPNQOPEVFCNQIFINNEWHDAVSRRKTEFTVNPSTGEVICOVAGSKEDV 67
 DB 1 AVPAIPNQOPEVFCNQIFINNEWHDAVSRRKTEFTVNPSTGEVICOVAGSKEDV 60

QY 421 ANNSTYGLAAAVTKDLKANYLSQALQAGTVWVNCYDVFRAGSPGGYKMSGSGRELGE 480
DB 439 ANNKYGLAAAVTKDLKANYLSQALQAGTVWVNCYDVFRAGSPGGYKMSGSGRELGE 498
QY 481 YGLQAYTEVKTVTKVPQKN 500
DB 499 YGLQAYTEVKTVTKVPQKN 518

RESULT 15

US-09-791-537-130904
; Sequence 130904, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 13095
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 130904
; LENGTH: 519
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-791-537-130904

Query Match 94.5%; Score 2480; DB 21; Length 519;
Best Local Similarity 94.2%; Pred. No. 2.8e-247;
Matches 471; Conservative 13; Mismatches 16; Indels 0; Gaps 0;

QY 1 MSRAATQAVPAPNQPEVFCNQIFINNEWHDAVSRKTPPTVNPSTGCVICQVAGCKEDV 60
DB 19 LSAAATSAVAPAHQPEVFCNQIFINNEWHDAVSRKTPPTVNPSTGCVICQVAGCKEDV 78
QY 61 DKAFEPGPAPQAGSPRRPMDASHGRLNRLADLIERDPTYLAALETLDNCKPYVSYL 120
DB 79 DKAVKAARAFAQAGSPRRMDASDRGRLLYRLADLIERDPTYLAALETLDNCKPYVSYL 138
QY 121 VLDLWVLKCLRYAGWADYHGKTIPIDGDFFSYTRHEPVGCGQIIPWNEPFLMQAWKL 180
DB 139 VLDLWVLKCLRYAGWADYHGKTIPIDGDFFSYTRHEPVGCGQIIPWNEPFLMQAWKL 198
QY 181 GPALATCNVVMKVAEQTPITALYVANLIKEAGFPGVVNI VPGFGPTAGAAIASHEVDV 240
DB 199 GPALATCNVVMKVAEQTPITALYVANLIKEAGFPGVVNI VPGFGPTAGAAIASHEVDV 258
QY 241 KVAFTGSTEIGRVIQVAAGSSNIIKRVITLGGKSPNII MSADMIWAVEQAHFALFTNQ 300
DB 259 KVAFTGSTEIGRVIQVAAGSSNIIKRVITLGGKSPNII MSADMIWAVEQAHFALFTNQ 318
QY 301 QCCACGSRFTVGEDIYDFEVSVAFAKSPVWNEFDSKIEGPGVDETFQPKILGYINT 360
DB 319 QCCACGSRFTVGEDIYDFEVSVAFAKSPVWNEFDSKIEGPGVDETFQPKILGYINT 378
QY 361 GKQEGAKLLCGGFIADRGYFIQPTVFGVQPGMTIAKEEIPGPMQILKEPTIEEVVGR 420
DB 379 GKQEGAKLLCGGFIADRGYFIQPTVFGVQPGMTIAKEEIPGPMQILKEPTIEEVVGR 438
QY 421 ANNSTYGLAAAVTKDLKANYLSQALQAGTVWVNCYDVFRAGSPGGYKMSGSGRELGE 480
DB 439 ANNKYGLAAAVTKDLKANYLSQALQAGTVWVNCYDVFRAGSPGGYKMSGSGRELGE 498
QY 481 YGLQAYTEVKTVTKVPQKN 500
DB 499 YGLQAYTEVKTVTKVPQKN 518

Search completed: June 24, 2003, 10:30:38
Job time : 183.017 secs

GenDate version 5.1.6
Copyright (c) 1995 - 2003 Computer 104

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:14:15 ; Search time 54.1289 Seconds
(without alignments)
2375 712 Million cell updates/ser

Title: US-09-830-751-4
Perfect score: 2625
Sequence: 1 MSAATQAVPAPNQPEVFC YGLQAYTEVKTIVTKVQKN 500

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1171708 seqs, 457144465 residues

Total number of hits satisfying chosen parameters: 1171708

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_New.*

- 1: /cgn2_6/ptodata/2/paa/PT_NTW_COMB.pep.*
- 2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
- 3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
- 4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
- 5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
- 6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
- 7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2625	100.0	500	5	US-09-830-751-4
2	2572	98.0	517	6	US-10-219-051B-6971
3	2372	98.0	517	6	US-10-219-051B-6975
4	2372	98.0	520	5	US-09-949-016-10653
5	2546.5	97.0	514	7	US-60-453-135-14871
6	2546.5	97.0	514	7	US-60-453-050-14871
7	2546.5	97.0	514	7	US-60-466-412-14871
8	2500	95.2	519	6	US-10-219-051B-6969
9	2500	95.2	519	6	US-10-219-051B-6973
10	1794	68.3	518	6	US-10-144-779-472
11	1791	68.2	532	5	US-60-449-616-9207
12	1790	68.2	518	5	US-09-724-676-64510
13	1790	68.2	518	5	US-09-724-676A-64510
14	1774	67.6	501	5	US-09-724-676-64517
15	1774	67.6	501	5	US-09-724-676A-64517
16	1769	67.4	500	6	US-10-219-051B-12834
17	1769	67.4	500	6	US-10-219-051B-12838
18	1756.5	66.9	516	5	US-09-724-676-64518
19	1756.5	66.9	516	5	US-09-724-676A-64518
20	1756.5	66.9	516	5	US-09-724-676A-64518
21	1756.5	66.9	516	5	US-09-724-676A-64519
22	1718	65.4	512	1	PCR US02-32371-2
23	1718	65.4	512	6	US-10-268-518-2
24	1716	65.4	501	6	US-10-316-253-18
25	1711	65.2	500	6	US-10-219-051B-12842
26	1711	65.2	500	6	US-10-219-051B-12836

27 1703.5 64.9 544 6 US-10-369-493-5715 Sequence 5715, Ap
28 1700 64.8 479 7 US-60-453-135-13160 Sequence 13160, A
29 1700 64.8 479 7 US-60-453-050-13160 Sequence 13160, A
30 1700 64.8 479 7 US-60-466-412-13160 Sequence 13160, A
31 1566.5 59.7 362 5 US-09-724-676-65471 Sequence 65471, A
32 1566.5 59.7 362 5 US-09-724-676A-65471 Sequence 65471, A
33 1543 58.8 549 6 US-10-219-051B-64510 Sequence 20, Appl
34 1543 58.8 549 6 US-10-425-114-49963 Sequence 49963, A
35 1543 58.8 552 6 US-10-425-114-49963 Sequence 49963, A
36 1538 58.6 464 5 US-09-724-676A-65684 Sequence 65684, A
37 1538 58.6 464 5 US-09-724-676A-65684 Sequence 65684, A
38 1535 58.5 552 6 US-10-425-114-59042 Sequence 59042, A
39 1521 57.9 550 6 US-10-438-246-8835 Sequence 8835, Ap
40 1518 57.8 539 6 US-10-424-599-264414 Sequence 264414, A
41 1515.5 57.7 540 6 US-10-424-599-106412 Sequence 196412, A
42 1512.5 57.6 439 5 US-09-949-016-9559 Sequence 9559, Ap
43 1512.5 57.6 439 5 US-60-453-135-13161 Sequence 13161, A
44 1512.5 57.6 439 7 US-60-453-050-13161 Sequence 13161, A
45 1512.5 57.6 439 7 US-60-466-412-13161 Sequence 13161, A

ALIGNMENTS

RESULT 1

US-09-830-751-4
: Sequence 4, Application US/09830751
: GENERAL INFORMATION:
: APPLICANT: Suthers, Patrick P.
: APPLICANT: Cameron, Douglas C.
: TITLE OF INVENTION: Production of 3-Hydroxypropionic Acid in Recombinant
: FILE REFERENCE: 940296 96617
: CURRENT APPLICANT: US/09830, 751
: CURRENT FILING DATE: 2000-08-30
: PRIOR APPLICANT: 65/151,440
: PRIOR FILING DATE: 1999-08-30
: PRIOR APPLICATION NUMBER: PCT/US00/23878
: PTO FILING DATE: 2000-08-30
: NIMRP OF SEQ ID NOS: 23
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO: 4
: LENGTH: 500
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-830-751-4

Query Match 100.0%; Score 2625; DB 5; Length 500;
Best Local Similarity 100.0%; Ident No. 736/229;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSAATQAVPAPNQPEVFCNQIFINNEHDAVSRKTFPTVNPSTGEVLCQVAESIKEDV 60
|||||
DB 1 MSAATQAVPAPNQPEVFCNQIFINNEHDAVSRKTFPTVNPSTGEVLCQVAESIKEDV 60
|||||
QY 61 DKARGGRRGAGQIGSPWPRMASHSOPILNPLADI.TERPNTYLALETLNKPKPVYSTL 120
|||||
DB 62 EKAREGEFGAFQLGSEFWPRMASHSRL.NPLAF.LFPGRITYLALETLNKPKPVYSTL 120
|||||
QY 101 VLLQMWLKECPYYAGWAFYHGTITPTGGEFFYTPHEFVGWVGGQLPWNEPILMQAWKL 180
|||||
DB 121 VOLDMVLCPLYYAGWAFYHGTITPTGGEFFYTPHEFVGWVGGQLPWNEPILMQAWKL 180
|||||
QY 181 GFALATGNNVMKVAEQTTTALYVANI.IKEA.FPTGVVHIVIP.TPTAA.VASHEWY 240
|||||
DB 181 GPALATGNNVMKVAEQTTTALYVANI.IKEAGFPVGVNIVPGEGPTAGAA.TASHEWD 240
|||||
QY 241 KVAFTGSTEIGFVQVAAGSSNLKPKPTLELCRKNLIMSQA.MWAFQAHFA.FFNQ 400
|||||
DB 241 KVAFTGSTEIGFVQVAAGSSNLKPKPTLELCRKNLIMSQA.MWAFQAHFA.FFNQ 400
|||||
QY 301 QCCAGSKTFQEDLY.EFVVF.SVAPAF.SWV.NP.TUSKTEGJLVLELQEKLLSYLNI 360
|||||

Db 301 OCCACSRFTVQEDIIYDEFVVRVARAKSRVGNCPDSTKTEGQPOVDETQPKKILGYINT 360
 QY 361 GKQBAKLLGGGTAARIGYFTQPTVFRILVQDGMTIAKEEIPSPVMQILKFKTIEEVVGR 420
 Db 361 GKQBAKLLGGGTAARIGYFTQPTVFRILVQDGMTIAKEEIPSPVMQILKFKTIEEVVGR 420
 QY 421 ANNSTYGLAAAVFTKLDKANYLSQALQAGTVMVNCYDVGAGSPGGYKMSGSGRELGE 480
 Db 421 ANNSTYGLAAAVFTKLDKANYLSQALQAGTVMVNCYDVGAGSPGGYKMSGSGRELGE 480
 QY 481 YGLQAYTEVKTIVKVPQKN 500
 Db 481 YGLQAYTEVKTIVKVPQKN 500

RESULT 2

US-10-219-051B-6971
 : Sequence 6971, Application US/10219051B
 : GENERAL INFORMATION:
 : APPLICANT: The General Hospital Corporation doing business as Massachusetts General
 : APPLICANT: Hospital / Bayer AG
 : TITLE OF INVENTION: Nucleotide sequences involved in pain
 : FILE REFERENCE: Lea 35693 Foreign Countries
 : CURRENT FILING DATE: 2003-05-09
 : PRIOR APPLICATION NUMBER: US 60/312,147
 : PRIOR FILING DATE: 2001-08-14
 : PRIOR APPLICATION NUMBER: US 60/346,382
 : PRIOR FILING DATE: 2001-11-01
 : PRIOR APPLICATION NUMBER: US 60/333,347
 : PRIOR FILING DATE: 2001-11-26
 : NUMBER OF SEQ ID NOS: 14715
 : SOFTWARE: Perl script
 : SEQ ID NO 6971
 : LENGTH: 517
 : TYPE: PRT
 : ORGANISM: Homo sapiens
 : PUBLICATION INFORMATION:
 : DATABASE ACCESSION NUMBER: SWISS-Prot / P05091
 : DATABASE ENTRY DATE: 2002-06-15
 US-10-219-051B-6971

Query Match: 98.0%; Score 2572; DB 6; Length 517;
 Best Local Similarity 98.2%; Pctd. No. 4,96,224;
 Matches 491; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
 QY 1 MSAAATQAVPAPNQOEVEFCNQIFINNEWHDAVSRKTPPTVNPSTGEVICOVAEGDKEDV 60
 Db 17 LSAAATQAVPAPNQOEVEFCNQIFINNEWHDAVSRKTPPTVNPSTGEVICOVAEGDKEDV 76
 QY 61 DKAREGRPGAFOLGSPWRMRDASHGRLLNRLADLIERDRTYLAALLETLDNGKPVVISYL 120
 Db 77 DKAVKAARAAPFOLGSPWRMRDASHGRLLNRLADLIERDRTYLAALLETLDNGKPVVISYL 136
 QY 121 VLDLDMVLKCLRYVAGWADKYHCKTIPIDGDFSYTRHEPVGVCGOIIPWNEPLLMOAWKL 180
 Db 137 VLDLDMVLKCLRYVAGWADKYHCKTIPIDGDFSYTRHEPVGVCGOIIPWNEPLLMOAWKL 196
 QY 181 GPALATGNVVMKVAEQTPLTALYVANLKEAGFPFGVNVIVPGPGTAGAAIASHEIWD 240
 Db 197 GPALATGNVVMKVAEQTPLTALYVANLKEAGFPFGVNVIVPGPGTAGAAIASHEIWD 256
 QY 241 KVAFTGSTELGRVIOVAAGSSNLKRVITPLEGKSPNITMSADMDWAVEQAHFALFFNQG 300
 Db 257 KVAFTGSTELGRVIOVAAGSSNLKRVITPLEGKSPNITMSADMDWAVEQAHFALFFNQG 316
 QY 301 QCCACSRFTVQEDIIYDEFVVRVARAKSRVGNCPDSTKTEGQPOVDETQPKKILGYINT 360
 Db 317 QCCACSRFTVQEDIIYDEFVVRVARAKSRVGNCPDSTKTEGQPOVDETQPKKILGYINT 376
 QY 361 GKQBAKLLGGGTAARIGYFTQPTVFRILVQDGMTIAKEEIPSPVMQILKFKTIEEVVGR 420
 Db 377 GKQBAKLLGGGTAARIGYFTQPTVFRILVQDGMTIAKEEIPSPVMQILKFKTIEEVVGR 436

QY 421 ANNSTYGLAAAVFTKLDKANYLSQALQAGTVMVNCYDVGAGSPGGYKMSGSGRELGE 480
 Db 437 ANNSTYGLAAAVFTKLDKANYLSQALQAGTVMVNCYDVGAGSPGGYKMSGSGRELGE 496
 QY 481 YGLQAYTEVKTIVKVPQKN 500
 Db 497 YGLQAYTEVKTIVKVPQKN 516

RESULT 3

US-10-219-051B-6975
 : Sequence 6975, Application US/10219051B
 : GENERAL INFORMATION:
 : APPLICANT: The General Hospital Corporation doing business as Massachusetts General
 : APPLICANT: Hospital / Bayer AG
 : TITLE OF INVENTION: Nucleotide sequences involved in pain
 : FILE REFERENCE: Lea 35693 Foreign Countries
 : CURRENT FILING DATE: 2003-05-09
 : PRIOR APPLICATION NUMBER: US 60/412,147
 : PRIOR FILING DATE: 2001-08-14
 : PRIOR APPLICATION NUMBER: US 60/446,482
 : PRIOR FILING DATE: 2001-11-01
 : PRIOR APPLICATION NUMBER: US 60/433,447
 : PRIOR FILING DATE: 2001-11-26
 : NUMBER OF SEQ ID NOS: 14715
 : SOFTWARE: Perl script
 : SEQ ID NO 6975
 : LENGTH: 517
 : TYPE: PRT
 : ORGANISM: Homo sapiens
 : PUBLICATION INFORMATION:
 : DATABASE ACCESSION NUMBER: SWISS-Prot / P05091
 : DATABASE ENTRY DATE: 2002-06-15
 US-10-219-051B-6975

Query Match: 98.0%; Score 2572; DB 6; Length 517;
 Best Local Similarity 98.2%; Pctd. No. 4,96,224;
 Matches 491; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
 QY 1 MSAAATQAVPAPNQOEVEFCNQIFINNEWHDAVSRKTPPTVNPSTGEVICOVAEGDKEDV 60
 Db 17 LSAAATQAVPAPNQOEVEFCNQIFINNEWHDAVSRKTPPTVNPSTGEVICOVAEGDKEDV 76
 QY 61 DKAREGRPGAFOLGSPWRMRDASHGRLLNRLADLIERDRTYLAALLETLDNGKPVVISYL 120
 Db 77 DKAVKAARAAPFOLGSPWRMRDASHGRLLNRLADLIERDRTYLAALLETLDNGKPVVISYL 136
 QY 121 VLDLDMVLKCLRYVAGWADKYHCKTIPIDGDFSYTRHEPVGVCGOIIPWNEPLLMOAWKL 180
 Db 137 VLDLDMVLKCLRYVAGWADKYHCKTIPIDGDFSYTRHEPVGVCGOIIPWNEPLLMOAWKL 196
 QY 181 GPALATGNVVMKVAEQTPLTALYVANLKEAGFPFGVNVIVPGPGTAGAAIASHEIWD 240
 Db 197 GPALATGNVVMKVAEQTPLTALYVANLKEAGFPFGVNVIVPGPGTAGAAIASHEIWD 256
 QY 241 KVAFTGSTELGRVIOVAAGSSNLKRVITPLEGKSPNITMSADMDWAVEQAHFALFFNQG 300
 Db 257 KVAFTGSTELGRVIOVAAGSSNLKRVITPLEGKSPNITMSADMDWAVEQAHFALFFNQG 316
 QY 301 QCCACSRFTVQEDIIYDEFVVRVARAKSRVGNCPDSTKTEGQPOVDETQPKKILGYINT 360
 Db 317 QCCACSRFTVQEDIIYDEFVVRVARAKSRVGNCPDSTKTEGQPOVDETQPKKILGYINT 376
 QY 361 GKQBAKLLGGGTAARIGYFTQPTVFRILVQDGMTIAKEEIPSPVMQILKFKTIEEVVGR 420
 Db 377 GKQBAKLLGGGTAARIGYFTQPTVFRILVQDGMTIAKEEIPSPVMQILKFKTIEEVVGR 436
 QY 421 ANNSTYGLAAAVFTKLDKANYLSQALQAGTVMVNCYDVGAGSPGGYKMSGSGRELGE 480
 Db 437 ANNSTYGLAAAVFTKLDKANYLSQALQAGTVMVNCYDVGAGSPGGYKMSGSGRELGE 496

```

QY 481 YGLOAYTEVKTIVTKVPQKN 500
Db 497 YGLOAYTEVKTIVTKVPQKN 516

RESULT 4
US-09-949-016-10653
; Sequence 10653, Application US/09949016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIORITY FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10653
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10653

```

```

Query Match      98.08, Score 2572, DB 5, Length 520,
Best Local Similarity 98.24, Pred. No. 4,90-224,
Matches 491, Conservative 2, Mismatches 7, Indels 0, Gaps 0.

QY 1 MSAAATQAVPAPNQPEVFCNQIFINNEHDAVSRKTPPTVNPSTGEVTCQVAGDKEDV 60
Db 20 LSAATQAVPAPNQPEVFCNQIFINNEHDAVSRKTPPTVNPSTGEVTCQVAGDKEDV 79
QY 61 DKAEPGPGAPFQIGSPWPPMDASHGSPLENPLADLIERDPTYLAAETLDNGKPVVISYL 120
Db 80 DKAVKAAPAAAFQIGSPWPPMDASHGSPLENPLADLIERDPTYLAAETLDNGKPVVISYL 139
QY 121 VLDLMLKCLRYAGWADKYHGKTIPIDGDFSYTHKEPVGCGQIIPWNPFLMQAKL 180
Db 140 VLDLMLKCLRYAGWADKYHGKTIPIDGDFSYTHKEPVGCGQIIPWNPFLMQAKL 199
QY 181 GPALATGNVVMKVAEQTPTLTYLVANLKEAGFPVGVNIVPGFGPTAGAAIASHEDVD 240
Db 200 GPALATGNVVMKVAEQTPTLTYLVANLKEAGFPVGVNIVPGFGPTAGAAIASHEDVD 259
QY 241 KVAFTGSTEIGRVIOVAAAGSSNLKRVTLLEGGKSPNIIMSDADMDWAVEQAHFALFFNOG 300
Db 260 KVAFTGSTEIGRVIOVAAAGSSNLKRVTLLEGGKSPNIIMSDADMDWAVEQAHFALFFNOG 319
QY 301 QCCAGSKPTVQEDLYDEFPVRSVARAKSPVGNPNPDSKTEGQPVDETFQFKILGYINT 360
Db 320 QCCAGSKPTVQEDLYDEFPVRSVARAKSPVGNPNPDSKTEGQPVDETFQFKILGYINT 379
QY 361 GKUGAKLLCGGGIAADRGYFIQPTVFGDVGQMTIAKEEIFSPVMQILKFKTIEVVVGR 420
Db 380 GKUGAKLLCGGGIAADRGYFIQPTVFGDVGQMTIAKEEIFSPVMQILKFKTIEVVVGR 449
QY 421 ANNSTYGLAAAVFTKLDKANYLSQALQAGIIVWNCYVFGAGSPFGYKMGSGRELGE 480
Db 440 ANNSTYGLAAAVFTKLDKANYLSQALQAGIIVWNCYVFGAGSPFGYKMGSGRELGE 499
QY 481 YGLOAYTEVKTIVTKVPQKN 500
Db 500 YGLOAYTEVKTIVTKVPQKN 519

RESULT 5
US-60-453-135-14871
; Sequence 14871, Application US/60453135

```

```

; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: IAKUBOWA, Olga
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001456
; CURRENT APPLICATION NUMBER: US/60/453,135
; PRIORITY FILING DATE: 2003-03-10
; NUMBER OF SEQ ID NOS: 82762
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14871
; LENGTH: 514
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-453-135-14871

Query Match      97.08, Score 2646, DB 7, Length 514,
Best Local Similarity 97.68, Pred. No. 9,90-222,
Matches 488, Conservative 2, Mismatches 7, Indels 3, Gaps 1.

QY 1 MSAAATQAVPAPNQPEVFCNQIFINNEHDAVSRKTPPTVNPSTGEVTCQVAGDKEDV 60
Db 17 LSAATQAVPAPNQPEVFCNQIFINNEHDAVSRKTPPTVNPSTGEVTCQVAGDKEDV 76
QY 61 DKAEPGPGAPFQIGSPWPPMDASHGSPLENPLADLIERDPTYLAAETLDNGKPVVISYL 120
Db 77 DKAVKAARAAAFQIGSPWPPMDASHGSPLENPLADLIERDPTYLAAETLDNGKPVVISYL 136
QY 121 VLDLMLKCLRYAGWADKYHGKTIPIDGDFSYTHKEPVGCGQIIPWNPFLMQAKL 180
Db 137 VLDLMLKCLRYAGWADKYHGKTIPIDGDFSYTHKEPVGCGQIIPWNPFLMQAKL 195
QY 181 GPALATGNVVMKVAEQTPTLTYLVANLKEAGFPVGVNIVPGFGPTAGAAIASHEDVD 240
Db 197 GPALATGNVVMKVAEQTPTLTYLVANLKEAGFPVGVNIVPGFGPTAGAAIASHEDVD 256
QY 241 KVAFTGSTEIGRVIOVAAAGSSNLKRVTLLEGGKSPNIIMSEADMIWAVEQAHFALFFNOG 300
Db 257 KVAFTGSTEIGRVIOVAAAGSSNLKRVTLLEGGKSPNIIMSEADMIWAVEQAHFALFFNOG 316
QY 301 QCCAGSKPTVQEDLYDEFPVRSVARAKSPVGNPNPDSKTEGQPVDETFQFKILGYINT 360
Db 317 QCCAGSKPTVQEDLYDEFPVRSVARAKSPVGNPNPDSKTEGQPVDETFQFKILGYINT 376
QY 361 GKUGAKLLCGGGIAADRGYFIQPTVFGDVGQMTIAKEEIFSPVMQILKFKTIEVVVGR 420
Db 377 GKUGAKLLCGGGIAADRGYFIQPTVFGDVGQMTIAKEEIFSPVMQILKFKTIEVVVGR 436
QY 421 ANNSTYGLAAAVFTKLDKANYLSQALQAGIIVWNCYVFGAGSPFGYKMGSGRELGE 480
Db 437 ANNSTYGLAAAVFTKLDKANYLSQALQAGIIVWNCYVFGAGSPFGYKMGSGRELGE 496
QY 481 YGLOAYTEVKTIVTKVPQKN 500
Db 497 YGLOAYTEVKTIVTKVPQKN 513

RESULT 6
US-60-453-050-14871
; Sequence 14871, Application US/60453050
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: LUKE, May
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; SLEWIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001457
; CURRENT APPLICATION NUMBER: US/60/453,050
; PRIORITY FILING DATE: 2003-03-10
; NUMBER OF SEQ ID NOS: 82762
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14871
; LENGTH: 514
; TYPE: PRT

```

ORGANISM: Homo sapiens
US-60-466-412-14871

Query Match 97.0%; Score 2546.5; DB 7; Length 514;
Best Local Similarity 97.6%; Pred. No. 9.9e-222;
Matches 488; Conservative 2; Mismatches 7; Indels 3; Gaps 1;

```

QY 1 MSAAATQAVPAPNQPEVFCNQIFINNEWHDAVSRKTFPIVNPSTGEVICVAVAGDKEDV 60
DB 17 LSAAATQAVPAPNQPEVFCNQIFINNEWHDAVSRKTFPIVNPSTGEVICVAVAGDKEDV 76
QY 61 DKAREGRPGAFOLGSPWRPMDSHGRLLNRLADLIERDRTYLAALFTLDNGKPPVVSYL 120
DB 77 DKAVKAARAFOLGSPWRPMDSHGRLLNRLADLIERDRTYLAALFTLDNGKPPVVSYL 136
QY 121 VLDMLVKCLRYAGWADKYHGKTIPIIDGDFSYTRHPEVGVCGQIIPWNPFLLMQAWKL 180
DB 137 VLDMLVKCLRYAGWADKYHGKTIPIIDGDFSYTRHPEVGVCGQIIPWNPFLLMQAWKL 196
QY 181 GPALATGNVVMKVAQOTPLTALYVANLKEAGFPVGVNIVPGEGPTAGAAIASHEDVD 240
DB 197 GPALATGNVVMKVAQOTPLTALYVANLKEAGFPVGVNIVPGEGPTAGAAIASHEDVD 256
QY 241 KVAFTGSTEIGRVIOVAAAGSSNLKRVTLGLGKSPNII MSDADMDWAVEQAHFALFFNQ 300
DB 257 KVAFTGSTEIGRVIOVAAAGSSNLKRVTLGLGKSPNII MSDADMDWAVEQAHFALFFNQ 316
QY 301 QCCAGSRTFVQHDIDYDFEVSVAARAKSRVGNPFDSKTEGQPOVDFTQPKKILGYINT 360
DB 317 QCCAGSRTFVQHDIDYDFEVSVAARAKSRVGNPFDSKTEGQPOVDFTQPKKILGYINT 376
QY 361 GKQEGAKLLCGGGIAADPGYFTQPTVFGDVQDQMTIAKEEIRGPMQILKFKTIEEVVGR 420
DB 377 GKQEGAKLLCGGGIAADPGYFTQPTVFGDVQDQMTIAKEEIRGPMQILKFKTIEEVVGR 436
QY 421 ANNSTYGLAAAVFTKDLKANYLSQALQAGTVMWNCYIVFSAQSPFGYKMGSGRELGE 480
DB 437 ANNSTYGLAAAVFTKDLKANYLSQALQAGTVMWNCYIVFSAQSPFGYKMGSGRELGE 496
QY 481 YGLQAYTEVKTVTVKVPQKN 500
DB 497 YGLQAYTE---VTVKVPQKN 513

```

RESULT 7

US-60-466-412-14871
Sequence 14871, Application US/60466412

GENERAL INFORMATION:
APPLICANT: CARGILL, Michele
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: G1001466
CURRENT APPLICATION NUMBER: US/60/466,412
CURRENT FILING DATE: 2003-04-30
NUMBER OF SEQ ID NOS: 429241
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14871
LENGTH: 514
TYPE: PRT
ORGANISM: Homo sapiens

US-60-466-412-14871

Query Match 97.0%; Score 2546.5; DB 7; Length 514;
Best Local Similarity 97.6%; Pred. No. 9.9e-222;
Matches 488; Conservative 2; Mismatches 7; Indels 3; Gaps 1;

```

QY 1 MSAAATQAVPAPNQPEVFCNQIFINNEWHDAVSRKTFPIVNPSTGEVICVAVAGDKEDV 60
DB 17 LSAAATQAVPAPNQPEVFCNQIFINNEWHDAVSRKTFPIVNPSTGEVICVAVAGDKEDV 76
QY 61 DKAREGRPGAFOLGSPWRPMDSHGRLLNRLADLIERDRTYLAALFTLDNGKPPVVSYL 120

```

```

DB 77 DKAVKAARAFOLGSPWRPMDSHGRLLNRLADLIERDRTYLAALFTLDNGKPPVVSYL 136
QY 121 VLDMLVKCLRYAGWADKYHGKTIPIIDGDFSYTRHPEVGVCGQIIPWNPFLLMQAWKL 180
DB 137 VLDMLVKCLRYAGWADKYHGKTIPIIDGDFSYTRHPEVGVCGQIIPWNPFLLMQAWKL 196
QY 181 GPALATGNVVMKVAQOTPLTALYVANLKEAGFPVGVNIVPGEGPTAGAAIASHEDVD 240
DB 197 GPALATGNVVMKVAQOTPLTALYVANLKEAGFPVGVNIVPGEGPTAGAAIASHEDVD 256
QY 241 KVAFTGSTEIGRVIOVAAAGSSNLKRVTLGLGKSPNII MSDADMDWAVEQAHFALFFNQ 300
DB 257 KVAFTGSTEIGRVIOVAAAGSSNLKRVTLGLGKSPNII MSDADMDWAVEQAHFALFFNQ 316
QY 301 QCCAGSRTFVQHDIDYDFEVSVAARAKSRVGNPFDSKTEGQPOVDFTQPKKILGYINT 360
DB 317 QCCAGSRTFVQHDIDYDFEVSVAARAKSRVGNPFDSKTEGQPOVDFTQPKKILGYINT 376
QY 361 GKQEGAKLLCGGGIAADPGYFTQPTVFGDVQDQMTIAKEEIRGPMQILKFKTIEEVVGR 420
DB 377 GKQEGAKLLCGGGIAADPGYFTQPTVFGDVQDQMTIAKEEIRGPMQILKFKTIEEVVGR 436
QY 421 ANNSTYGLAAAVFTKDLKANYLSQALQAGTVMWNCYIVFSAQSPFGYKMGSGRELGE 480
DB 437 ANNSTYGLAAAVFTKDLKANYLSQALQAGTVMWNCYIVFSAQSPFGYKMGSGRELGE 496
QY 481 YGLQAYTEVKTVTVKVPQKN 500
DB 497 YGLQAYTE---VTVKVPQKN 513

```

RESULT 8

US-10-219-051B-6969

Sequence 6969, Application US/10219051B
GENERAL INFORMATION:
APPLICANT: The General Hospital Corporation doing business as Massachusetts General
TITLE OF INVENTION: Nucleotide sequences involved in pain
FILE REFERENCE: ICA 45693 Foreign Countries
CURRENT APPLICATION NUMBER: US/10/219,051B
CURRENT FILING DATE: 2003-05-09
PRIOR APPLICATION NUMBER: US 60/412,147
PRIOR FILING DATE: 2001-08-14
PRIOR APPLICATION NUMBER: US 60/446,482
PRIOR FILING DATE: 2001-11-01
PRIOR APPLICATION NUMBER: US 60/443,447
PRIOR FILING DATE: 2001-11-26
NUMBER OF SEQ ID NOS: 14715
SOFTWARE: Perl script
SEQ ID NO 6969
LENGTH: 519
TYPE: PRT
ORGANISM: Rattus norvegicus
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: SWISS-Prot / P11884
DATABASE ENTRY DATE: 2002-06-15

US-10-219-051B-6969

Query Match 95.2%; Score 2500; DB 6; Length 519;
Best Local Similarity 94.8%; Pred. No. 1.6e-217;
Matches 474; Conservative 12; Mismatches 17; Indels 0; Gaps 0;

```

QY 1 MSAAATQAVPAPNQPEVFCNQIFINNEWHDAVSRKTFPIVNPSTGEVICVAVAGDKEDV 60
DB 19 LSAAATQAVPAPNQPEVFCNQIFINNEWHDAVSRKTFPIVNPSTGEVICVAVAGDKEDV 78
QY 61 DKAREGRPGAFOLGSPWRPMDSHGRLLNRLADLIERDRTYLAALFTLDNGKPPVVSYL 120
DB 79 DKAVKAARAFOLGSPWRPMDSHGRLLNRLADLIERDRTYLAALFTLDNGKPPVVSYL 148
QY 121 VLDMLVKCLRYAGWADKYHGKTIPIIDGDFSYTRHPEVGVCGQIIPWNPFLLMQAWKL 180
DB 139 VLDMLVKCLRYAGWADKYHGKTIPIIDGDFSYTRHPEVGVCGQIIPWNPFLLMQAWKL 198

```

QY 181 GPALATGNNVVKVAEQTPLTALYVANIKEAGFPBGVNVIVPCFGPTACAAATASHEDVD 240
 Db 199 GPALATGNNVVKVAEQTPLTALYVANIKEAGFPBGVNVIVPCFGPTACAAATASHEDVD 258
 QY 241 KVAFTGSTEIGRVIVQAAGSSNLKRVITLGLGKSPNIIIMSDADMWAVEQAIFALFPNG 300
 Db 259 KVAFTGSTEIGRVIVQAAGSSNLKRVITLGLGKSPNIIIMSDADMWAVEQAIFALFPNG 318
 QY 301 QCCAGSKTFVQEDYDEFVVRVSVAKAKSVVGNPFDKSTEQGPQVDETOFKKILGYINT 360
 Db 319 QCCAGSKTFVQEDYDEFVVRVSVAKAKSVVGNPFDKSTEQGPQVDETOFKKILGYINT 378
 QY 361 GKQEGAKLGGGIAAARGYFIQPTVFGDVQDGMTIAKEBIFGPMQILKFKTIEBVGCR 420
 Db 379 GKQEGAKLGGGIAAARGYFIQPTVFGDVQDGMTIAKEBIFGPMQILKFKTIEBVGCR 438
 QY 421 ANNSTYGLAAAVFTKDLKANYLSQALQAGIVWVNYVVRSAUSPFSYKMSGSPETGE 480
 Db 439 ANNSTYGLAAAVFTKDLKANYLSQALQAGIVWVNYVVRSAUSPFSYKMSGSPETGE 498
 QY 481 YGLOAYTEVKTIVTKVPQKN 500
 Db 499 YGLOAYTEVKTIVTKVPQKN 518

RESULT 9

US-10-219-051B-6973
 ; Sequence 6973, Application us/10219051B
 ; GENERAL INFORMATION:
 ; APPLICANT: The General Hospital Corporation doing business as Massachusetts General
 ; APPLICANT: Hospital / Bayer AG
 ; TITLE OF INVENTION: Nucleotide sequences involved in pain
 ; FILE REFERENCE: Lea 35693 Foreign Countries
 ; CURRENT APPLICATION NUMBER: US/10/219,051B
 ; CURRENT FILING DATE: 2003-05-09
 ; PRIOR APPLICATION NUMBER: US 60/3312,147
 ; PRIOR FILING DATE: 2001-08-14
 ; PRIOR APPLICATION NUMBER: US 60/346,382
 ; PRIOR FILING DATE: 2001-11-31
 ; PRIOR APPLICATION NUMBER: US 60/333,347
 ; PRIOR FILING DATE: 2001-11-26
 ; NUMBER OF SEQ ID NOS: 14715
 ; SOFTWARE: Perl script
 ; SEQ ID NO 6973
 ; LENGTH: 519
 ; TYPE: PRT
 ; ORGANISM: Rattus norvegicus
 ; PUBLICATION INFORMATION:
 ; DATABASE ACCESSION NUMBER: SWISS-Prot / P11884
 ; DATABASE ENTRY DATE: 2002-06-15
 ; US-10-219-051B-6973

Query Match 95.2%, Score 2509, DB 6, Length 519,
 Best Local Similarity 94.8%, Pred. No. 1,6e-217;
 Matches 474; Conservative 12; Mismatches 14; Indels 0; Gaps 0;
 QY 1 MSAAATQAVAPNPOOPEVFCNQIFINNEHWDVSRKTIPTVNPSTGEVTCQVABGKHEIV 60
 Db 19 LSAATSAVAPNPOOPEVFCNQIFINNEHWDVSRKTIPTVNPSTGEVTCQVABGKHEIV 78
 QY 61 DKAREGPPAFGLSGPMPMDASHSGPLINPLADLIERDPTVLAALLETILNKKPVVISYL 120
 Db 79 EKAVKAAAGAFGLSGPMPMDASHSGPLINPLADLIERDPTVLAALLETILNKKPVVISYL 148
 QY 121 VLDLWLVKCLRYAGWADKYHGKTIPTVNPSTGEVTCQVABGKHEIV 180
 Db 139 VLDLWLVKCLRYAGWADKYHGKTIPTVNPSTGEVTCQVABGKHEIV 198
 QY 181 GPALATGNNVVKVAEQTPLTALYVANIKEAGFPBGVNVIVPCFGPTACAAATASHEDVD 240
 Db 199 GPALATGNNVVKVAEQTPLTALYVANIKEAGFPBGVNVIVPCFGPTACAAATASHEDVD 258

QY 241 KVAFTGSTEIGRVIVQAAGSSNLKRVITLGLGKSPNIIIMSDADMWAVEQAIFALFPNG 300
 Db 259 KVAFTGSTEIGRVIVQAAGSSNLKRVITLGLGKSPNIIIMSDADMWAVEQAIFALFPNG 318
 QY 301 QCCAGSKTFVQEDYDEFVVRVSVAKAKSVVGNPFDKSTEQGPQVDETOFKKILGYINT 360
 Db 319 QCCAGSKTFVQEDYDEFVVRVSVAKAKSVVGNPFDKSTEQGPQVDETOFKKILGYINT 378
 QY 361 GKQEGAKLGGGIAAARGYFIQPTVFGDVQDGMTIAKEBIFGPMQILKFKTIEBVGCR 420
 Db 379 GKQEGAKLGGGIAAARGYFIQPTVFGDVQDGMTIAKEBIFGPMQILKFKTIEBVGCR 438
 QY 421 ANNSTYGLAAAVFTKDLKANYLSQALQAGIVWVNYVVRSAUSPFSYKMSGSPETGE 480
 Db 439 ANNSTYGLAAAVFTKDLKANYLSQALQAGIVWVNYVVRSAUSPFSYKMSGSPETGE 498
 QY 481 YGLOAYTEVKTIVTKVPQKN 500
 Db 499 YGLOAYTEVKTIVTKVPQKN 518

RESULT 10

US-10-144-779-472
 ; Sequence 472, Application US/10144779
 ; GENERAL INFORMATION:
 ; APPLICANT: SUBRAMANIAN, Mani et al.
 ; TITLE OF INVENTION: MOUSE ORTHOLOGS OF HUMAN DISEASE GENES,
 ; TITLE OF INVENTION: PROTEINS ENCODED BY THESE MOUSE GENES, AND USES THEREOF
 ; FILE REFERENCE: C1001235
 ; CURRENT APPLICATION NUMBER: US/10/144,779
 ; CURRENT FILING DATE: 2002-05-15
 ; NUMBER OF SEQ ID NOS: 864
 ; SOFTWARE: PastSeq for Windows Version 4.0
 ; SEQ ID NO 472
 ; LENGTH: 518
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 ; US-10-144-779-472

Query Match 94.4%, Score 1794, DB 6, Length 518,
 Best Local Similarity 66.0%, Pred. No. 1,6e-153;
 Matches 328; Conservative 77; Mismatches 92; Indels 0; Gaps 0;
 QY 4 AATQAVAPNPOOPEVFCNQIFINNEHWDVSRKTIPTVNPSTGEVTCQVABGKHEIV 63
 Db 21 ASLQILPSPTNIFTKYTKTIPINNPNQNSGSEGVFPVNPATGEQVDEQAHKVVIDKA 80
 QY 64 PEKPPGAFGLSGPMPMDASHSGPLINPLADLIERDPTVLAALLETILNKKPVVISYL 123
 Db 81 VQAAFLAFSLGSPMPMDASHSGPLINPLADLIERDPTVLAALLETILNKKPVVISYL 140
 QY 124 DMVVKCLRYAGWADKYHGKTIPTVNPSTGEVTCQVABGKHEIV 183
 Db 141 QGVKILRYAGWADKYHGKTIPTVNPSTGEVTCQVABGKHEIV 200
 QY 184 LATQNVVVMKVAEQTPLTALYVANIKEAGFPBGVNVIVPCFGPTACAAATASHEDVD 243
 Db 201 LQCNVIVVMKVAEQTPLTALYVANIKEAGFPBGVNVIVPCFGPTACAAATASHEDVD 260
 QY 244 PTGSETGPTVQAAGSSNLKRVITLGLGKSPNIIIMSDADMWAVEQAIFALFPNG 303
 Db 261 PTGSETGPTVQAAGSSNLKRVITLGLGKSPNIIIMSDADMWAVEQAIFALFPNG 320
 QY 404 CAGSKTFVQEDYDEFVVRVSVAKAKSVVGNPFDKSTEQGPQVDETOFKKILGYINT 463
 Db 421 TAGSKTFEESLYDEFVVRVSVAKAKSVVGNPFDKSTEQGPQVDETOFKKILGYINT 480
 QY 464 BKAKLGGGIAAARGYFIQPTVFGDVQDGMTIAKEBIFGPMQILKFKTIEBVGCR 523
 Db 481 PGAKLGGGIAAARGYFIQPTVFGDVQDGMTIAKEBIFGPMQILKFKTIEBVGCR 540
 QY 424 STYGLAAAVFTKDLKANYLSQALQAGIVWVNYVVRSAUSPFSYKMSGSPETGE 483
 Db 441 STYGLAAAVFTKDLKANYLSQALQAGIVWVNYVVRSAUSPFSYKMSGSPETGE 500

```

Db      441  SUEGLVAAVFTNDINKALMVSSAMQAGTVWJINCYNALNAQSPRRPRPKMS:NGCFMGEFGL 500
QY      484  QAYTEVKTIVTKVPQKN 500
       : :|:|||||:|:|:|
Db      501  REYSEVKTIVTKVPQKN 517

RESULT 11
US-09-949-016-q207
; Sequence 9207, Application US/09949016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9207
; LENGTH: 532
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9207

Query Match      68.2%, Score 1791, DB 5, Length 532;
Best Local Similarity 66.2%, Pred. No. 3, le-153;
Matches 329; Conservative 75; Mismatches 93; Indels 0; Gaps 0;

QY      4  AATQAVPAPNQOPEVFNQIPIINNEHDAVSRKTEPTVNPSTGEVTCQVAGDKEDVDKA 63
Db      35  ASLHLPSPTNLNLEIKTKYKPIINNEWONSSEGRVFVYNFATGEQVCEVQEAADKADIDKA 94

QY      64  REGRRPGAFQLGSPRRMDASHGRLNRLADLERORTYLAALLETLDNCKPVIISLYIDL 123
Db      95  VQAARLAFLSGSVRRMDASFERGRLDLKLDLVERDRAVLATNESLNGGKPFLOQPYVDL 154

QY      124  DMVLKCLRYAGWADKYHGKTIPIDGDDFFSYTRHEPVGVCQIIPWNFFLLMQANKLGA 183
Db      155  QGVIKTRFYAGWADKYHGWTIPVDGYDFTTRHEPIGVCGQIIPWNFFLLMFANKIAPA 214

QY      184  LATGNVNVVMKYAEOTPLTALYVNLIKEAGFPQGVNIVPGFPTAGAAIAHSHEDVDKA 243
Db      215  LCCGNTVVKPFAEQIPIISALYMGALIKEAGFPQGVINILPGYGTAGAAIAHSHIGIDKTA 274

QY      244  FTGSTEIGRVIQVAAGSSNLKRVITELGKSPNIIIMSDMDWAVBQAHFALFFNQGC 303
Db      275  FTGSTEVGKLIQEAAGRSNLKRVITELGKSPNIIIFADADLYAVQAHUQGVFFNQGC 334

QY      304  CAGSTFTVQEDIIYDFVVRVARAKSRVGNPNFDSKTEQGPQVDETQFKKILCYINTGK 363
Db      335  TAGSRIFVEESIYDFVVRVERAKRWVSGSPDPTTEQGPQIDKKQYKNLLELIQSGVA 394

QY      364  EGAKLLCGGLGIAADRGYFIQFTVFGDVGQCMIIAKEEIFGPMUILLKPKTIEEVGKRAN 423
Db      395  EGAKLECGKGLGRKRGFFIEPTVESNVTDDMRITAKEEIEGPMUILLKPKTMDVETERAN 454

QY      424  STYGLAAAVFTKDLIDKANYLISQALQAGTVWNCYIVFGAQSPPGGYKMSGSGREIGEYGL 483
Db      455  SDFGLVAAVFTNDINKALTVSSAMQAGTVWJINCYNALNAQSPRRPRPKMS:NGCFMGEFGL 514

QY      484  QAYTEVKTIVTKVPQKN 500
Db      515  REYSEVKTIVTKVPQKN 531

```



```

Query Match      68.2%, Score 1790, DB 5, Length 518.
Best Local Similarity 66.0%, Prod. No. 4, he-154;
Matches 328; Conservative 76; Mismatches 93; Indels 0; Gaps 0;

QY 4 AATQAVPAPNQQPEVFCNOIFINNEWHDAVSRKTFPTVNPSTGEVICOVARGQKEDVDKA 63
Db 21 ASHLIPSPITNLEIKYTKIFINNEWQNSGRVFFVYNFATGQVCEVGEADKADIDKA 80
QY 64 PEAPPAFALGSSWPMPLASHSSLLNPLADLIEEDRTYLAALETLGNKPKYVSYLVLD 123
Db 81 VQAAAPLAFSGSWPMRMDASGRILLDKLADLVERDAVLATMESLNGKPKPFQAFVPL 140
QY 124 DMVLKCLRYAGWADKYHCKTIPIDGDFSYTRHEPVGVCQJLIIPWNEPILLMQAWKLGPA 183
Db 141 QGVKTFERYAGWAKIRHMTIPVDGYTFTRHFPFUSVQJLIIPWNEPILLMQAWKLTAPA 200
QY 184 LATGNVVMKVAEQPTPLTALXVANLIKKEAGPPGCVNIVPGFGPTAGAAIASHEDVDKVA 243
Db 201 LCGNTVVIKPAEQPTPLSALYMGALIKFACFPCCVINILPGYCGPTAGAAIASHIGIDKTA 260
QY 244 FTGSTEIRVIOVAAGSSNLKRVTLIEIGSKSPNIIIMSDAMDMWAVEQAHFALFENQGCC 303
Db 261 FTGSTEVRGKLIOEAAAGRSNLKRVTLIEIGSKSPNIIIFADADLDYAVEQAHQGVTFNQGCC 320
QY 304 CAGSRTFVGEEDYDEFVVSVARAKSRVVGVPFSKTEGCGPQVLETFQPKKILGYINTGKO 363
Db 321 TACSRTFVFEESYDEFVVSVERAKKPIVRSPTFTTEGCGPQIDKKYNNKILEIGSSVA 380
QY 364 EGAKLLCGSSIAAIDYTFIQPTVFIQWJQJMIIAKEELIEFGVMMILKEKTEIEFVVSFKANN 423
Db 381 EGAKLFCGSGKIGRKGFFIPTVFSNVTDMRIAKEELIEFGVQELIEFKTMDVIERANN 440
QY 424 STYGAAAVFTKDLKANYLSGALQAGTVWVNYVFTAGCGFFSGSYKMSGRPELSEYGL 483
Db 441 SDGGLVAAVFTINDKALTVSAMQAGTVWVNYVFTAGCGFFSGSYKMSGRPELSEYGL 500
QY 484 QAYTEVKTIVTKVPOKN 500
Db 501 REYSEVKTIVTKIPOKN 517

RESULT 14
US-09-724-676-64517
; Sequence 64517, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724.676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 64517
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-64517

Query Match      67.6%, Score 1774, DB 5, Length 501;
Best Local Similarity 67.0%, Prod. No. 9, 70-152;
Matches 335; Conservative 67; Mismatches 98; Indels 0; Gaps 0;

QY 1 MSAAATQAVPAPNQQPEVFCNOIFINNEWHDAVSRKTFPTVNPSTGEVICOVARGQKEDV 60
Db 1 MSSSGTPELPLVLLTDLKIYTKIFINNEWHDSVSKKPPVNPATREIICQVERGKEDV 60
QY 61 DKAPRGPACAFOLGSPWPMPLASHSSLLNPLADLIEEDPTTYLAALFTLNGKPKYVSYL 120
Db 61 DKAVRAARCAFOLGSPWMTMLASRPSRLYLKLAULIEERDPLLLATMESLNGSKLYSNAYL 120
QY 121 VLDLDMVILKCLIRYAGWADKYHCKTIPIDGDFSYTRHEPVGVCQJLIIPWNEPILLMQAWK 180
Db 121 NLGALIKLIPYAGWAKIRHMTIPVDGYTFTRHFPFUSVQJLIIPWNEPILLMQAWKLTAPA 180
QY 181 CPALATGNVVMKVAEQPTPLTALXVANLIKKEAGPTVWVRIIVPGFGPTAGAAIASHEDVD 240
Db 181 GPALSCGNTVVVKPAEQPTPLTALHVASLIKKEAGPPGCVNIVPGYGTAGAAISSHMDID 240
QY 241 KVAFPGSTIEGHWLVAAGSSNLKRVTLIEIGSKSPNIIIMSLAIMWAVEQAHFALFENQ 300
Db 241 KVAFPGSTIEVRGKLIEAAGKSNLKRVTLIEIGSKSPCIVLADADLDNAVERFAHGVYHQ 300
QY 301 LCGNTVVIKPAEQPTPLSALYMGALIKFACFPCCVINILPGYCGPTAGAAIASHIGIDKTA 360
Db 301 QCCIAASPIFVFEESYDEFVVSVERAKKPIVRSPTFTTEGCGPQIDKKYNNKILEIGSSVA 360
QY 361 EGAKLLCGSSIAAIDYTFIQPTVFIQWJQJMIIAKEELIEFGVMMILKEKTEIEFVVSFK 420
Db 361 EGAKLLCGSSIAAIDYTFIQPTVFIQWJQJMIIAKEELIEFGVMMILKEKTEIEFVVSFK 420
QY 421 ANNSTYGLAAAVFTKDLKANYLSGALQAGTVWVNYVFTAGCGFFSGSYKMSGRPELSEYGL 480
Db 421 ANNSTYGLAAAVFTKDLKANYLSGALQAGTVWVNYVFTAGCGFFSGSYKMSGRPELSEYGL 480
QY 481 YGLQAYTEVKTIVTKVPOKN 500
Db 481 YGFHEYTEVKTIVTKISOKN 500

RESULT 15
US-09-724-676A-64517
; Sequence 64517, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724.676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 64517
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-64517

Query Match      67.6%, Score 1774, DB 5, Length 501;
Best Local Similarity 67.0%, Prod. No. 9, 70-152;
Matches 335; Conservative 67; Mismatches 98; Indels 0; Gaps 0;

QY 1 MSAAATQAVPAPNQQPEVFCNOIFINNEWHDAVSRKTFPTVNPSTGEVICOVARGQKEDV 60
Db 1 MSSSGTPELPLVLLTDLKIYTKIFINNEWHDSVSKKPPVNPATREIICQVERGKEDV 60
QY 61 DKAPRGPACAFOLGSPWPMPLASHSSLLNPLADLIEEDPTTYLAALFTLNGKPKYVSYL 120
Db 61 DKAVRAARCAFOLGSPWMTMLASRPSRLYLKLAULIEERDPLLLATMESLNGSKLYSNAYL 120
QY 121 VLDLDMVILKCLIRYAGWADKYHCKTIPIDGDFSYTRHEPVGVCQJLIIPWNEPILLMQAWK 180
Db 121 NLGALIKLIPYAGWAKIRHMTIPVDGYTFTRHFPFUSVQJLIIPWNEPILLMQAWKLTAPA 180
QY 181 CPALATGNVVMKVAEQPTPLTALXVANLIKKEAGPTVWVRIIVPGFGPTAGAAIASHEDVD 240
Db 181 GPALSCGNTVVVKPAEQPTPLTALHVASLIKKEAGPPGCVNIVPGYGTAGAAISSHMDID 240
QY 241 KVAFPGSTIEGHWLVAAGSSNLKRVTLIEIGSKSPNIIIMSLAIMWAVEQAHFALFENQ 300
Db 241 KVAFPGSTIEVRGKLIEAAGKSNLKRVTLIEIGSKSPCIVLADADLDNAVERFAHGVYHQ 300
QY 301 LCGNTVVIKPAEQPTPLSALYMGALIKFACFPCCVINILPGYCGPTAGAAIASHIGIDKTA 360
Db 301 QCCIAASPIFVFEESYDEFVVSVERAKKPIVRSPTFTTEGCGPQIDKKYNNKILEIGSSVA 360
QY 361 EGAKLLCGSSIAAIDYTFIQPTVFIQWJQJMIIAKEELIEFGVMMILKEKTEIEFVVSFK 420

```

```

Db      361  GKKEGAKIFCGGGRWGNKGYFVQPTVFSNVTDEMRTAKEEIEGPPQQIMKFKSLDDVIKR 420
QY      421  ANNSTYGLAAAVFTKDLKANYLSQALQACTVWVWNCYDVFAGOSPPGGYKMSGSGRELGE 480
Db      421  ANNTFYGLSAGVETKDIDKAITISSALQACTVWVWNCYGVVSACPPGGFKMSGNGRELGE 480
QY      481  YGLQAYTEVKTVTVKVPKN 500
Db      481  YGFHEYTEVKTVTVKISOKN 500

```

Search completed: June 24, 2003, 10:35:21
 Job time : 56.1289 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 GenCore Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:10:05 ; Search time 19.164 seconds
(without alignments)
2512.114 Million cell updates/sec

Title: US-09-830-751-4
Perfect score: 242
Sequence: 1 MSAATCAVAPAPQAPQPEVFC YRLQAVTEVKTIVKVPQFN 500

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2572	98.0	517	1 DEH0E2	aldehyde d-hydroge
2	2500	95.2	519	1 S33564	aldehyde d-hydroge
3	2480	94.5	519	1 148966	aldehyde d-hydroge
4	2436	92.8	520	1 S09030	aldehyde d-hydroge
5	2434	92.7	530	1 S00364	aldehyde d-hydroge
6	1939	73.9	517	1 A40872	aldehyde d-hydroge
7	1802	64.4	504	1 S14629	aldehyde d-hydroge
8	1794	64.4	494	2 S74224	aldehyde d-hydroge
9	1774	67.6	501	1 DEH0E1	aldehyde d-hydroge
10	1763	67.2	501	2 S14752	aldehyde d-hydroge
11	1757	66.9	501	2 J05553	aldehyde d-hydroge
12	1744	66.4	501	1 J01004	aldehyde d-hydroge
13	1743	66.4	501	2 J04524	aldehyde d-hydroge
14	1736	66.1	500	1 S02302	aldehyde d-hydroge
15	1718	65.4	512	1 A55684	aldehyde d-hydroge
16	1716	65.4	501	1 A32616	aldehyde d-hydroge
17	1703.5	64.9	544	2 D88449	Protein F54D8 x li
18	1587	60.5	497	2 J04924	aldehyde d-hydroge
19	1543	58.8	538	2 T06483	aldehyde d-hydroge
20	1539	58.6	542	2 T02361	aldehyde d-hydroge
21	1537	58.6	549	2 T04983	rf2 nuclear restor
22	1499	57.1	496	1 A46725	omega-crystallin
23	1498.5	57.1	514	2 D86472	hypothetical prote
24	1425.5	54.3	497	1 A59055	aldehyde d-hydroge
25	1422.5	54.2	511	1 S31308	aldehyde d-hydroge
26	1378.5	52.5	503	2 T54216	aldehyde d-hydroge
27	1354	51.6	498	2 C83717	NADP-dependent ald
28	1325	50.5	495	2 B60614	aldehyde d-hydroge
29	1305.5	49.7	496	2 S43114	aldehyde d-hydroge

ALIGNMENTS

RESULT 1

DEH0E2

N-Alternative names: aldehyde dehydrogenase E2; aldehyde dehydrogenase I
C:Species: Homo sapiens (man)
C:Date: 31-Mar-1988 sequence revision 17-Nov-1995 #rev: change 03-Jun-2002
C:Accession: A29975, S00804, A23503, A27509, A26743, 139432, 139433, 139430
R:Hsu, L.C.; Bengel, P.F.; Yoshida, A
Genomics 2, 57-65, 1988
A:Title: Genomic structure of the human mitochondrial aldehyde dehydrogenase gene.
A:Reference number: A29975, M010.88256152, PMID:2838413
A:Accession: A29975
A:Molecule type: DNA
A:Residues: 1-517 <HSU>
A:Cross-references: GR M26760, NID:917897, PIRN AAA51694 1; PID:917898
R:Braun, T.; Heber, E.; Singh, S.; Agarwal, D.P.; Goedde, H.W.
FEBS Lett. 233, 440, 1988
A:Reference number: S00804
A:Accession: S00804
A:Molecule type: mRNA
A:Residues: 1-6, 'AFA', 10, 'F', 13, 49, 'DEA-'
A:Cross-references: EMBL:X55409, NID:928605
A:Note: correction to A26743
R:Bengel, J.; Kaiser, P.; Joernvall, H.
Eur. J. Biochem. 153, 13-28, 1985
A:Title: Mitochondrial aldehyde dehydrogenase from human liver. Primary structure, di
A:Reference number: A23503, M010.8605846, PMID:4065146
A:Accession: A23503
A:Molecule type: protein
A:Residues: 'A', 18-517, 'HEM'
A:Note: the sequence shown is presumably that of the mature protein, however, the num
R:Braun, T.; Heber, E.; Singh, S.; Agarwal, D.P.; Goedde, H.W.
Nucleic Acids Res. 15, 3179, 1987
A:Title: Isolation and sequence analysis of a full length cDNA clone coding for human
A:Reference number: A27509, M010.8714840, PMID:663350
A:Accession: A27509
A:Molecule type: mRNA
A:Residues: 1-6, 'AWFAWA', 10, 'F', 12, 'VS', 15, 'FHFGF', 21, 27-79, 'FEGPG', 86-336, 'V', 338-5
A:Cross-references: EMBL:X04004, NID:928605
R:Braun, T.; Heber, E.; Singh, S.; Agarwal, D.P.; Goedde, H.W.
FEBS Lett. 215, 233-236, 1987
A:Title: Evidence for a signal peptide at the amino-terminal end of human mitochondri
A:Reference number: A26743, M010.8721061, PMID:3582651
A:Accession: A26743
A:Molecule type: mRNA
A:Residues: 1-6, 'AWFAWA', 10, 'F', 12, 'VS', 15, 'FHFGF', 21, 27-79, 'FEGPG', 86-336, 'V', 338-5
A:Cross-references: EMBL:X04004, NID:928605
R:Hsu, L.C.; Tai, K.; Fujiyoshi, T.; Kuriachi, K.; Yoshida, A.
Proc. Natl. Acad. Sci. U.S.A. 82, 3771-3775, 1985
A:Title: Cloning of cDNAs for human aldehyde dehydrogenases 1 and 2.
A:Reference number: 139432, M010.85216574, PMID:2987944

A:Accession: J39432
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 119-336, 'V', 338-517 <S02>
 A:Cross-references: GB:K03001, NID:q178495, P1IN:AAK54500.1; P1D:q178495
 A:Note: thirty-three tryptic peptides were also sequenced
 R:Yoshida, A.; Ikawa, M.; Hsu, L.C.; Tani, K.
 Alcohol 2, 103-106, 1985
 A:Title: Molecular abnormality and cDNA cloning of human aldehyde dehydrogenases.
 A:Reference number: J39431; PMID:85252089; PMID:4015823
 A:Accession: J39433
 A:Status: preliminary, translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 119-336, 'V', 338-517 <S02>
 A:Cross-references: GB:M26760, NID:q178397, P1IN:AA51634.1, P1D:q178398
 R:Agarwal, D.P.; Goede, H.W.
 Isozymes Curr Top Biol Med Res 16, 21-48, 1987
 A:Title: Human aldehyde dehydrogenase isozymes and alcohol sensitivity.
 A:Reference number: J39430; PMID:87279033; PMID:3610592
 A:Accession: J39430
 A:Status: preliminary, translated from GB/EMBL/DDBJ
 A:Molecule type: mRNA
 A:Residues: 214-215, 'S', 217, 'R', 219, 246, 'F', 248, 336, 'V', 338, 425, 'E', 426, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 843, 844, 845, 846, 847, 848, 849, 850, 851, 852, 853, 854, 855, 856, 857, 858, 859, 860, 861, 862, 863, 864, 865, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, 885, 886, 887, 888, 889, 890, 891, 892, 893, 894, 895, 896, 897, 898, 899, 900, 901, 902, 903, 904, 905, 906, 907, 908, 909, 910, 911, 912, 913, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923, 924, 925, 926, 927, 928, 929, 930, 931, 932, 933, 934, 935, 936, 937, 938, 939, 940, 941, 942, 943, 944, 945, 946, 947, 948, 949, 950, 951, 952, 953, 954, 955, 956, 957, 958, 959, 960, 961, 962, 963, 964, 965, 966, 967, 968, 969, 970, 971, 972, 973, 974, 975, 976, 977, 978, 979, 980, 981, 982, 983, 984, 985, 986, 987, 988, 989, 990, 991, 992, 993, 994, 995, 996, 997, 998, 999, 1000
 A:Note: the sequence is misidentified as aldehyde dehydrogenase 1
 C:Genetics:
 A:Gene: GDB:ALDH2
 A:Cross-references: GDB:119668; OMIM:100650
 A:Map position: 12q24.2-12q24.2
 A:Introns: 39/4; 74/4; 120/4; 147/2; 184/3; 227/3; 265/3; 300/1; 361/2; 416/3; 450/2; 500/1
 C:Complex: homotetramer
 C:Function:
 A:Description: catalyzes the oxidation of an aldehyde to an acid using NAD+ and water
 A:Pathway: ethanol catabolism
 A:Note: enzymes with this activity are involved in diverse metabolic pathways in various C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology C:Keywords: alcohol metabolism; homotetramer; liver; mitochondrion; NAD; oxidoreductase F:1-17/Domain: transit peptide (mitochondrion) #status predicted <SIC>
 F:18-517/Product: aldehyde dehydrogenase (NAD+) 2 #status experimental <MAT>
 F:75-335/Domain: aldehyde dehydrogenase homology <ALDH>
 F:211-291/Domain: NAD binding #status predicted <NAD>
 F:285/Active site: Glu #status predicted
 F:319/Active site: Cys #status experimental
 F:472/Binding site: NAD (Cys) #status predicted

Query Match 98.0%; Score 2572; DB 1; Length 517;
 Best Local Similarity 98.2%; Pred. No. 9e-191;
 Matches 491; Conservative 7; Indels 0; Gaps 0;

QY 1 MSAAATGAPAPNQQPEVFCNQLFINNEHDAVSRKTEPTVNPSTGEVICOVAGGKEDV 60
 DB 17 LSAATGAPAPNQQPEVFCNQLFINNEHDAVSRKTEPTVNPSTGEVICOVAGGKEDV 76
 QY 61 DKAREGRPGAFQLCSPWRMDASHGRGLNRLADLIERDRTVLALETLDNGKPYVSYL 120
 DB 77 DKAVKAARAAPQGLSPWRMDASHGRGLNRLADLIERDRTVLALETLDNGKPYVSYL 136
 QY 121 VLEDMVLKCLRYAGWADYHCKTIPIDGDFSYTRHEEVGVCGIIPWNPFLLMQAWKL 180
 DB 137 VLDLMVLKCLRYAGWADYHCKTIPIDGDFSYTRHEEVGVCGIIPWNPFLLMQAWKL 196
 QY 181 GPALATGNVVMKVBOTPLTALYVANLILKEAGPPGVNIVPGFGPTAGAAATASHEDVD 240
 DB 197 GPALATGNVVMKVBOTPLTALYVANLILKEAGPPGVNIVPGFGPTAGAAATASHEDVD 256
 QY 241 KVAFTGSTEIGRVIOVAAASSNKPVTLIELGCKSPNTIMSDADIMWAVEQAHAFPTNQC 300
 DB 257 KVAFTGSTEIGRVIOVAAAGSSNLKRVILELGCKSNILMSADIMWAVEQAHAFPTNQC 316
 QY 301 QCCACGSRFTVQEDIDYDEFWERSVARAKSRVGVNIPFUSKTEGSGQVDETQPKKILGYINT 360
 DB 317 QCCACGSRFTVQEDIDYDEFWERSVARAKSRVGVNIPFUSKTEGSGQVDETQPKKILGYINT 376

QY 361 GKQFGAKLLCGGGIAADRGYFTQPTVFGVQDQMTAKKEIPQVWQILKKTIEVWGR 420
 DB 377 GKQFGAKLLCGGGIAADRGYFTQPTVFGVQDQMTAKKEIPQVWQILKKTIEVWGR 436
 QY 421 ANNSTVGLAAAVHFKLQKANYLSUALAGLVWVNYGVVFACTPFCYKMSGSKFELGE 480
 DB 437 ANNSTVGLAAAVHFKLQKANYLSUALAGLVWVNYGVVFACTPFCYKMSGSKFELGE 496
 QY 481 YGLQAYTEVKTIVTKVPOKN 500
 DB 497 YGLQAYTEVKTIVTKVPOKN 516

RESULT 2

S03564
 aldehyde dehydrogenase (NAD) (EC 1.2.1.3) 2 precursor, mitochondrial rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 28-Feb-1990 #sequence revision 31 Jan 1997 #text change 03 Jun 2002
 C:Accession: S03564; A:27713; S12993; S17492
 R:Farres, J.; Guan, K.L.; Weiner, H.
 Eur. J. Biochem. 180, 67-74, 1989
 A:Title: Primary structures of rat and bovine liver mitochondrial aldehyde dehydrogenase
 A:Reference number: S03564; PMID:2540003
 A:Accession: S03564
 A:Molecule type: mRNA
 A:Residues: 1-519 <FAR>
 A:Cross-references: EMBL:X14977; NID:q55604; P1IN:CAA3401.1; P1D:q55605
 A:Experimental source: Strain Sprague-Dawley; liver
 R:Farres, J.; Guan, K.L.; Weiner, H.
 Biochem Biophys Res Commun 150, 1083-1087, 1988
 A:Title: Sequence of the signal peptide for rat liver mitochondrial aldehyde dehydrogenase
 A:Reference number: A27713; PMID:88134217; PMID:442060
 A:Accession: A27713
 A:Molecule type: mRNA
 A:Residues: 1-29 <FAR>
 A:Cross-references: GR:M19030; NID:q202847; P1IN:AAA40719.1; P1D:q202848
 R:Diwan, J.J.; Pallwal, R.; Kaftan, E.; Bawa, R.
 FEBS Lett. 273, 215-218, 1990
 A:Title: A mitochondrial protein fraction catalyzing transport of the K(+) analog Tl(+) into the mitochondrion
 A:Reference number: S12993; PMID:91042184; PMID:1699808
 A:Accession: S12993
 A:Molecule type: protein
 A:Residues: 327-340 <DIW>
 R:Jeng, J.; Weiner, H.
 Arch. Biochem. Biophys. 289, 214-222, 1991
 A:Title: Purification and characterization of catalytically active precursor of rat 1
 A:Reference number: S17492; PMID:91378548; PMID:1898068
 A:Accession: S17492
 A:Molecule type: protein
 A:Residues: 1-19 <JEN>
 C:Genetics:
 A:Genome: nuclear
 C:Complex: homotetramer
 C:Function:
 A:Description: catalyzes the oxidation of an aldehyde to an acid using NAD+ and water
 A:Pathway: ethanol catabolism
 A:Note: enzymes with this activity are involved in diverse metabolic pathways in vari C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology C:Keywords: alcohol metabolism; homotetramer; liver; mitochondrion; NAD; oxidoreducta F:1-19/Domain: transit peptide (mitochondrion) #status predicted <TNC>
 F:20-519/Product: aldehyde dehydrogenase (NAD+) 2 #status predicted <MAT>
 F:77-341/Domain: aldehyde dehydrogenase homology <ALDH>
 F:287,321/Active site: Glu, Cys #status predicted
 F:474/Binding site: NAD (Cys) #status predicted

Query Match 95.2%; Score 2660; DB 1; Length 519;
 Best Local Similarity 94.8%; Pred. No. 3.4e-185;
 Matches 474; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

QY 1 MSAAATGAPAPNQQPEVFCNQLFINNEHDAVSRKTEPTVNPSTGEVICOVAGGKEDV 60
 DB 19 LSAATGAPAPNQQPEVFCNQLFINNEHDAVSRKTEPTVNPSTGEVICOVAGGKEDV 76

A:Residues: 92-520 <EAP>
 A:Experimental source: liver
 R:Lee, J.E.; Cho, Y.D.
 Biochem. Biophys. Res. Commun. 189, 450-454, 1992
 A:Title: Purification and characterization of bovine brain gamma-aminobutyraldehyde dehydrogenase
 A:Reference number: PQ0543; MUID:93080596; PMID:1449496
 A:Accession: PQ0543
 A:Molecule type: protein
 A:Residues: 22-34 <LEE>
 A:Experimental source: brain
 C:Comment: This enzyme may be responsible for gamma aminobutyraldehyde dehydrogenase activity
 C:Complex: homotetramer
 C:Function:
 A:Description: catalyzes the oxidation of an aldehyde to an acid using NAD+ and water
 A:Pathway: ethanol catabolism
 A:Note: enzymes with this activity are involved in diverse metabolic pathways in various
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 C:Keywords: alcohol metabolism; homotetramer; liver; mitochondrion; NAD; oxidoreductase
 F:1-21/Domain: transit peptide (mitochondrion) #status predicted <SIG>
 F:22-520/Product: aldehyde dehydrogenase (NAD+) 2 #status predicted <MAI>
 F:78-342/Domain: aldehyde dehydrogenase homology <ALDD>
 F:288-322/Domain: aldehyde dehydrogenase homology <ALDD>
 F:288-322/Active site: Glu, Cys #status predicted
 F:475/Binding site: NAD (Cys) #status predicted

Query Match 92.88; Score 2436; DB 1; Length 520;
 Best Local Similarity 91.88; Pred. No 2 90-180;
 Matches 457; Conservative 21; Mismatches 20; Indels 0; Gaps 0;

```

QY 3 AAATQAVAPNQOPEVFNQIFINNEWHDAVSRKFTPTVNPSTGEVICOVAGKEDVDK 62
DB 22 SAATQAVPTNQOPEVLYNQIFINNEWHDAVSRKFTPTVNPSTGDVICHVABGKADVD 81
QY 63 AREGPGCAFOLGSPWRMDASHGRLLNRLADLIERDRTYLAALLETDLNGKPYVLSYLD 122
DB 82 AVKARAFAFOLGSPWRMDASERGLNRLADLIERDRTYLAALLETDLNGKPYVLSYLD 141
QY 123 LDWVLKCLRYAGWADKYHGKTIPIIDGDFSYTRHEPVGVCQIIIPWNPFLMLQAWKLG 182
DB 142 LDWVLKCLRYAGWADKYHGKTIPIIDGDFSYTRHEPVGVCQIIIPWNPFLMLQAWKLG 201
QY 183 ALATGNVVMKVAEQTPLTALYVANLIKEAGPPGVNIVPGFGPTAGAAATASHEDVDK 242
DB 202 ALATGNVVMKVAEQTPLTALYVANLIKEAGPPGVNIVPGFGPTAGAAATASHEDVDK 261
QY 243 AFTGSTIEGRVIVAAAGSSNKRVTLELGGKSPNIIIMSDUADWAVEQAHAFFALFENQGC 302
DB 262 AFTGSTIEGRVIVAAAGSSNKRVTLELGGKSPNIIIMSDUADWAVEQAHAFFALFENQGC 321
QY 303 CCAGSRTEFVEDIYAEFVRSVARAKSRVGNPNPDSKTQGGQVDETOFKKILGYINTCK 362
DB 322 CCAGSRTEFVEDIYAEFVRSVARAKSRVGNPNPDSKTQGGQVDETOFKKILGYINTCK 381
QY 363 QERAKLLCGGRTAADRGYFIQPTVFGVDQGMTLAKEEIFGPMQILKFKTIEEVVGRAN 422
DB 382 QERAKLLCGGRTAADRGYFIQPTVFGVDQGMTLAKEEIFGPMQILKFKTIEEVVGRAN 441
QY 423 NSTYGLAAAVFTKDLKANYLSQALQAGTVVNCYDVFQAQSPFGGYKMSGSGRELGEYG 482
DB 442 NSTYGLAAAVFTKDLKANYLSQALQAGTVVNCYDVFQAQSPFGGYKMSGSGRELGEYG 501
QY 483 GLQAYTEVKTVTVKVPQKN 500
DB 502 GLQAYTEVKTVTVKVPQKN 519

```

RESULT 5
 S00364
 aldehyde dehydrogenase (NAD) (EC 1.2.1.3) 2, mitochondrial horse (tentative sequence)
 C:Species: Equus caballus (domestic horse)
 C:Date: 30-Jun-1989 #sequence_revision 31-Jan-1997 #text_change 04-Jun-2002
 R:Johansson, J.; von Bahr-Lindstrom, H.; Jeck, R.; Woenckhaus, C.; Joernvall, H.
 Eur. J. Biochem. 172, 527-533, 1988

A:Title: Mitochondrial aldehyde dehydrogenase from horse liver: Characterizations of the S
 A:Reference number: S00364; MUID:88166740; PMID:3450012
 A:Accession: S00364
 A:Molecule type: protein
 A:Residues: 1-500 <JOH>
 A:Note: amino terminal residue is uncertain, 1 Ser and 1 Leu were also found; forms h
 C:Complex: homotetramer
 C:Function:
 A:Description: catalyzes the oxidation of an aldehyde to an acid using NAD+ and water
 A:Pathway: ethanol catabolism
 A:Note: enzymes with this activity are involved in diverse metabolic pathways in vari
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 C:Keywords: alcohol metabolism; homotetramer; liver; mitochondrion; NAD; oxidoreducta
 F:58-322/Domain: aldehyde dehydrogenase homology <ALDD>
 F:288-402/Active site: Glu, Cys #status predicted
 F:455/Binding site: NAD (Cys) #status predicted

Query Match 92.79; Score 2414; DB 1; Length 500;
 Best Local Similarity 92.68; Pred. No. 3 90-180;
 Matches 452; Conservative 22; Mismatches 20; Indels 0; Gaps 0;

```

QY 2 SAATQAVAPNQOPEVFNQIFINNEWHDAVSRKFTPTVNPSTGEVICOVAGKEDVDK 61
DB 1 AAATQAVAPNQOPEVFNQIFINNEWHDAVSRKFTPTVNPSTGEVICOVAGKEDVDK 60
QY 62 KAPGPGCAFOLGSPWRMDASHGRLLNRLADLIERDRTYLAALLETDLNGKPYVLSYLD 121
DB 61 KAPGPGCAFOLGSPWRMDASHGRLLNRLADLIERDRTYLAALLETDLNGKPYVLSYLD 120
QY 122 LDWVLKCLRYAGWADKYHGKTIPIIDGDFSYTRHEPVGVCQIIIPWNPFLMLQAWKLG 181
DB 121 LDWVLKCLRYAGWADKYHGKTIPIIDGDFSYTRHEPVGVCQIIIPWNPFLMLQAWKLG 180
QY 182 PALATGNVVMKVAEQTPLTALYVANLIKEAGPPGVNIVPGFGPTAGAAATASHEDVDK 241
DB 181 PALATGNVVMKVAEQTPLTALYVANLIKEAGPPGVNIVPGFGPTAGAAATASHEDVDK 240
QY 242 VAPTGSTIEGRVIVAAAGSSNKRVTLELGGKSPNIIIMSDUADWAVEQAHAFFALFENQGC 301
DB 241 VAPTGSTIEGRVIVAAAGSSNKRVTLELGGKSPNIIIMSDUADWAVEQAHAFFALFENQGC 300
QY 302 QERAKLLCGGRTAADRGYFIQPTVFGVDQGMTLAKEEIFGPMQILKFKTIEEVVGRAN 361
DB 301 CCAGSRTEFVEDIYAEFVRSVARAKSRVGNPNPDSKTQGGQVDETOFKKILGYINTCK 360
QY 362 KQERAKLLCGGRTAADRGYFIQPTVFGVDQGMTLAKEEIFGPMQILKFKTIEEVVGRAN 421
DB 361 KQERAKLLCGGRTAADRGYFIQPTVFGVDQGMTLAKEEIFGPMQILKFKTIEEVVGRAN 420
QY 422 NSTYGLAAAVFTKDLKANYLSQALQAGTVVNCYDVFQAQSPFGGYKMSGSGRELGEYG 481
DB 421 NSTYGLAAAVFTKDLKANYLSQALQAGTVVNCYDVFQAQSPFGGYKMSGSGRELGEYG 480
QY 482 GLQAYTEVKTVTVKVPQKN 500
DB 481 GLQAYTEVKTVTVKVPQKN 499

```

RESULT 6
 A0872
 aldehyde dehydrogenase (NAD) (EC 1.2.1.3) 5 precursor, mitochondrial human
 C:Species: Homo sapiens (man)
 C:Date: 27-Mar-1992 #sequence_revision 4 Jan 1997 #text_change 03-Jun-2002
 C:Accession: A0872
 R:Hsu, L.C.; Chang, W.C.
 J. Biol. Chem. 266, 12267-12269, 1991
 A:Title: Cloning and characterization of a new tunnel human aldehyde dehydrogenase
 A:Reference number: A0872; MUID:91286241; PMID:2061411
 A:Accession: A0872
 A:Molecule type: DNA
 A:Residues: 1-517 <HSU>
 A:Cross-references: GB:M63967; NID:q437184; P10N:AAA46840.1; P1D:q1264008
 A:Note: 86-Ala and 107-Leu were also found

C:Comment: Based on similarity, this form is assumed to be mitochondrial and to have pro

C:Genetics:

A:Gene: GDB:ALDH5; ALDHX

A:Cross-references: GDB:128788

A:Map position: 9p13-9p13

A:Note: the only info occurs before the initiator codon

C:Complex: homotetramer (probably)

C:Function:

A:Description: catalyzes the oxidation of an aldehyde to an acid using NAD⁺ and water

A:Pathway: ethanol catabolism

A:Note: enzymes with this activity are involved in diverse metabolic pathways in vari

C:Superfamily: aldehyde dehydrogenase (NAD⁺); aldehyde dehydrogenase homology

C:Keywords: alcohol metabolism; cytosol; homotetramer; liver; NAD; oxidoreductase

F:1-20/Domain: transit peptide (mitochondrial) #status predicted <N>

F:21-517/Product: aldehyde dehydrogenase (NAD⁺) #status predicted <NAT>

F:75-339/Domain: aldehyde dehydrogenase homology <ALDH>

F:285,311/Active site: Glu, Cys #status predicted

Query Match 73.9%, Score 1939; DB 1; Length 517;
Best Local Similarity 72.9%; Pred. No. 7,2e-142;
Matches 364; Conservative 61; Mismatches 72; Indels 2; Gaps 1;

QY 2 SAATQAPAPNQOPEVCNOIFINNEWHDAVSRTPTVPSTGEVICOVAGGKEDVD 61
DB 20 SAA--ALPSPILAPDPYNLFNNEWDQAVSKTPTVNTTGEVGHVAGDRAVD 77
QY 62 KAKGPPAPVJGSPWPMFMDASHSGRLNRLADLTERDPTVLALETLDNGKPYVIVLD 121
DB 78 PAVKAAKRVFPLGSPWRPMDASERGLNRLADLVERDYLASLETLDNGKPYFORSYAL 137
QY 122 DLDMVLKCLRYAGWADKYHCKTIPIDGDFSYTRHPRVGVCGQIIPWNPFLMOAWKLG 181
DB 138 DLDEVIKVRVYFAGWAKWHGKTIPMDGQHFCTKHPEPVGVCGQIIPWNPFLMOQWKL 197
QY 182 PALATGNVVMKVAEOTPLTALYVANLIKKEAGPPGVNVVPGPPTAGAAIASHEDVDK 241
DB 198 PALATGNVVMKVAEOTPLTALYVANLIKKEAGPPGVNVVPGPPTAGAAIASHEDVDK 257
QY 242 VAFSTGSLGRVIAVAASNSNLRKVTLELGGKSPNIMSLALMIMWAVEQAHPALFFNAG 301
DB 258 VAFSTGSLGVHLQKAA--TNSLKRVTLLELGGKSPNIVLAIAWGEIAVEPHEALFFNAG 317
QY 302 CCAASRTPVQDIYDERVWRSVAKRSVGVNPIPSKTPGQPVVIELEPKKIDGYINT 361
DB 318 CCAASRTPVQDIYDERVWRSVAKRSVGVNPIPSKTPGQPVVIELEPKKIDGYINT 377
QY 362 KQETAKLLGGGIAADWYPTQPTVPFSDVQGMPTIAKEEIPGVPMTLAKFKTIERVVGRA 421
DB 378 QKEGAKLLGGGERGEGFFIKPIVFGSVQDMMIAKEEIPGVPQPIKPKKIERVVEFA 447
QY 422 NNSTYGLAAAVFTKDLKANYLSQALQAGTVVWNCYDVFQAQSPFGGKMSGRELGEY 481
DB 438 NNTYGLAAAVFTKDLKANYLSQALQAGTVVWNTYNTVTCHTPFGGKESGRELGEY 497
QY 482 GLQAYTEVKTIVTKVPQKN 500
DB 498 GLQAYTEVKTIVTKVPQKN 516

RESULT 7

S14629
aldehyde dehydrogenase (NAD) (EC 1.2.1.3) 1, cytosolic - chicken
C:Species: Gallus gallus (chicken)
C:Date: 21-Nov-1993 #sequence_revision 31-Jan-1997 #text_change 63-Jun-2002
C:Accession: S14629
R:Godbout, R.
submitted to the EMBL Data Library, April 1991
A:Description: High levels of aldehyde dehydrogenase transcripts in the undifferentiated
A:Reference number: S14629
A:Accession: S14629
A:Molecule type: mRNA
A:Residues: 1-509 <Gp>
A:Cross-references: EMBL X58869, NID 93012, PIRN CAA41679.1, PIR P36303

A:Experimental source: retina
C:Complex: homotetramer

C:Function:

A:Description: catalyzes the oxidation of an aldehyde to an acid using NAD⁺ and water

A:Pathway: ethanol catabolism

A:Note: enzymes with this activity are involved in diverse metabolic pathways in vari

C:Superfamily: aldehyde dehydrogenase (NAD⁺); aldehyde dehydrogenase homology

C:Keywords: alcohol metabolism; cytosol; homotetramer; liver; NAD; oxidoreductase

F:57-331/Domain: aldehyde dehydrogenase homology <ALDH>

F:277,311/Active site: Glu, Cys #status predicted

F:464/Binding site: NAD (Cys) #status predicted

Query Match 68.6%, Score 1802; DB 1; Length 509;
Best Local Similarity 67.3%; Pred. No. 2,7e-131;
Matches 305; Conservative 67; Mismatches 96; Indels 0; Gaps 0;

QY 3 AAATCAVAPNQQEVEFCNIFINNEWHDAVSRTPTVPSTGEVICOVAGGKEDVD 62
DB 11 APVLPALPEPLKDLKIKYIKIFINNEWHDSVSKKEEVPNANLEKICEVAEGSKADIDK 70
QY 63 APGPPGAFOLGSPWRPMDASHSGRLNRLADLTERDPTVLALETLDNGKPYVIVLD 122
DB 71 AVKAAKRAFEIGSGPWPMDASERGLNRLADLVERDPLTATMEADGGKLESTAYLMD 130
QY 123 LDMLVKCLRYAGWADKYHCKTIPIDGDFSYTRHPRVGVCGQIIPWNPFLMOAWKLG 182
DB 141 LDMLVKCLRYAGWADKYHCKTIPIDGDFSYTRHPRVGVCGQIIPWNPFLMOAWKLG 190
QY 183 ALATGNVVMKVAEOTPLTALYVANLIKKEAGPPGVNVVPGPPTAGAAIASHEDVDK 242
DB 191 ALATGNVVMKVAEOTPLTALYVANLIKKEAGPPGVNVVPGPPTAGAAIASHEDVDK 250
QY 244 APTGSTRTPGTVIAAASNSNLRKVTLELGGKSPNIMSLALMIMWAVEQAHPALFFNAG 302
DB 251 SETSGTEVGLIKKEAGKTNLRKVTLELGGKSPNIPADADLDEAAFAHGLIYHGQGC 310
QY 404 CCAASRTPVQDIYDERVWRSVAKRSVGVNPIPSKTPGQPVVIELEPKKIDGYINT 462
DB 411 CCAASRTPVQDIYDERVWRSVAKRSVGVNPIPSKTPGQPVVIELEPKKIDGYINT 470
QY 464 QKEGAKLLGGGIAADWYPTQPTVPFSDVQGMPTIAKEEIPGVPMTLAKFKTIERVVGRA 422
DB 471 QKEGAKLLGGGERGEGFFIKPIVFGSVQDMMIAKEEIPGVPQPIKPKKIERVVEFA 440
QY 423 NNSTYGLAAAVFTKDLKANYLSQALQAGTVVWNCYDVFQAQSPFGGKMSGRELGEY 482
DB 431 NNTYGLAAAVFTKDLKANYLSQALQAGTVVWNCYDVFQAQSPFGGKMSGRELGEY 490
QY 483 GLQAYTEVKTIVTKVPQKN 500
DB 491 GLQAYTEVKTIVTKVPQKN 508

RESULT 8

S74224
aldehyde dehydrogenase (NAD) (EC 1.2.1.3) 2 - mouse
N:Alternate names: retinaldehyde-specific dehydrogenase
C:Species: Mus musculus (house mouse)
C:Date: 04-Dec-1997 #sequence_revision 12-Dec 1997 #text_change 63-Jun-2002
C:Accession: S74224
R:Zhao, D.; McCafferty, P.; Ivins, K.J.; Neve, R.L.; Hodan, P.; Chiu, W.W.; Draeger, U.
Eur. J. Biochem. 240, 15-22, 1996
A:Title: Molecular identification of a major retinoic-acid-synthesizing enzyme, a ret
A:Reference number: S74224; MIM:606985; PMJ:874840
C:Accession: S74224
A:Molecule type: mRNA
A:Residues: 1-499 <ZHA>
A:Cross-references: EMBL X74274, NID 9140868, PIRN CAA87665.1; PIR G1430869
A:Experimental source: strain C3H/He; cell type embryonal carcinoma; cell line p19 te
C:Genetics:
A:Gene: RALDH-2
C:Superfamily: aldehyde dehydrogenase (NAD⁺); aldehyde dehydrogenase homology
C:Keywords: NAD; oxidoreductase

Db 181 GPALSGNTVVVKPAEQTPLTALHVASLIKEAGFPFGVNVIPGCGPTAGAAISSHMDID 240
 QY 241 KVAFTGSTEIGRIVQVAGSSNLKRVTLLELGKSPNIIIMSDAIMWAVQAHFALFNQ 400
 Db 241 KVAFTGSTEIGRIVQVAGSSNLKRVTLLELGKSPNIIIMSDAIMWAVQAHFALFNQ 400
 QY 301 QCCAGSRTFQEDIEYDFVVRKSVARAKSKVRNPPFQSKTQKSPQVDETFQKILGYINT 360
 Db 301 QCCAGSRTFQEDIEYDFVVRKSVARAKSKVRNPPFQSKTQKSPQVDETFQKILGYINT 360
 QY 361 GKQEGAKLLCGGSIADGNYIFLQPTVFHVGWMTIAKEELFPHVMILKFKETIEEVVGR 420
 Db 361 GKQEGAKLLCGGSIADGNYIFLQPTVFHVGWMTIAKEELFPHVMILKFKETIEEVVGR 420
 QY 421 ANNSTYGLAAAVFTKDLKANYLSQALQAGTVMWNCYDVFCAQSPFGGKMSGSGRELGE 480
 Db 421 ANNSTYGLAAAVFTKDLKANYLSQALQAGTVMWNCYDVFCAQSPFGGKMSGSGRELGE 480
 QY 481 YGLQAYTEVKTIVTKVPKN 500
 Db 481 YGFHEYTEVKTIVTKISOKN 500

RESULT 10

S14752
 aldehyde dehydrogenase (NAD) (EC 1.2.1.3), cytosolic - sheep
 C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
 C:Date: 10-Mar-1997 #sequence_revision 13-Sep-1998 #text_change 03-Jun-2002
 C:Accession: S78582
 R: Tweedie, J.W.
 submitted to the EMBL Data Library, July 1994
 A:Reference number: S78582
 A:Accession: S78582
 A:Molecule type: mRNA
 A:Residues: 1-501
 A:Cross references: EMBL:U12761, NID:J527681, FIDN:AAA85435.1, IID:J527682
 A:Experimental source: liver
 R:Kitson, J.M.; Hill, J.P.; Midwinter, G.G.
 Biochem J 275, 207-210, 1991
 A:Title: Identification of a catalytically essential nucleophilic residue in sheep liver
 A:Reference number: S14752, MUID:91207293, FUID:2618476

A:Accession: S14752
 A:Residues: 298-308 <KIT>
 A:Experimental source: liver
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 C:Keywords: NAD; oxidoreductase
 F:59-323/Domain: aldehyde dehydrogenase homology <ALD>
 F:302/Active site: Cys #status experimental

Query Match 67.2%, Score 1763, DB 2, Length 501,
 Best Local Similarity 66.6%, Pred. No. 2.7e-128;
 Matches 333; Conservative 67; Mismatches 100; Indels 0; Gaps 0;

QY 1 MSAATAQVAPAPNOOPEVFCNQIFINNEHDAVSRKTFPTVNPSTGEVICQVAGSKEDV 60
 Db 1 MSSAMPDVPAPLTNLQFYTKIFINNEHSSVSKGKPPFNATEKLEVEEGKADV 60
 QY 61 DKAREGRPAFQSGPWRMDASHGRLNRLADLIERDRTYLAALETLNGKPYVISYL 120
 Db 61 DKAVKAAPQAFQSGPWRMDASERGLNKLADLIERDRTYLAALETLNGKPYVISYL 120
 QY 121 VLDLWVLCRYAGWADKYHGKTIPIGDGFFSTHREPVGVCQGIIPWNEPFLLMQAWKL 180
 Db 121 MGLGCTKTCRYAGWADKYHGKTIPIGDGFFSTHREPVGVCQGIIPWNEPFLLMQAWKL 180
 QY 181 GPALATGNVVMKVAEQTPLTALXVANLIKEAGFPFGVNVIPGCGPTAGAAISSHMDID 240
 Db 181 GPALSGNTVVVKPAEQTPLTALHVASLIKEAGFPFGVNVIPGCGPTAGAAISSHMDID 240
 QY 241 KVAFTGSTEIGRIVQVAGSSNLKRVTLLELGKSPNIIIMSDAIMWAVQAHFALFNQ 400
 Db 241 KVAFTGSTEIGRIVQVAGSSNLKRVTLLELGKSPNIIIMSDAIMWAVQAHFALFNQ 400

QY 301 QCCAGSRTFQEDIEYDFVVRKSVARAKSKVRNPPFQSKTQKSPQVDETFQKILGYINT 360
 Db 301 QCCAGSRTFQEDIEYDFVVRKSVARAKSKVRNPPFQSKTQKSPQVDETFQKILGYINT 360
 QY 361 GKQEGAKLLCGGSIADGNYIFLQPTVFHVGWMTIAKEELFPHVMILKFKETIEEVVGR 420
 Db 361 GKQEGAKLLCGGSIADGNYIFLQPTVFHVGWMTIAKEELFPHVMILKFKETIEEVVGR 420
 QY 421 ANNSTYGLAAAVFTKDLKANYLSQALQAGTVMWNCYDVFCAQSPFGGKMSGSGRELGE 480
 Db 421 ANNSTYGLAAAVFTKDLKANYLSQALQAGTVMWNCYDVFCAQSPFGGKMSGSGRELGE 480
 QY 481 YGLQAYTEVKTIVTKVPKN 500
 Db 481 YGFHEYTEVKTIVTKISOKN 500

RESULT 11

JC5553
 aldehyde dehydrogenase (NAD) (EC 1.2.1.3) - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change 03-Jun-2002
 C:Accession: JC5553
 R:Kathmann, E.C.; Lipsky, J.J.
 Biochem Biophys Res Commun 236, 627-631, 1997
 A:Title: Cloning of a cDNA encoding a constitutively expressed rat liver cytosolic al-
 A:Reference number: J05553, MUID:9182470, FUID:9240474
 A:Accession: JC5553
 A:Molecule type: mRNA
 A:Residues: 1-501 <KAT>
 A:Cross references: DDBJ:AF001806, NID:Q2181212, FIDN:AA053304.1, PID:G2183213
 A:Experimental source: liver
 C:Comment: This enzyme catalyzes the irreversible oxidation of a wide variety of alde-
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 C:Keywords: oxidoreductase
 F:59-323/Domain: aldehyde dehydrogenase homology <ALD>

Query Match 66.9%, Score 1757, DB 2, Length 501,
 Best Local Similarity 66.2%, Pred. No. 7.9e-128;
 Matches 331; Conservative 69; Mismatches 100; Indels 0; Gaps 0;

QY 1 MSAATAQVAPAPNOOPEVFCNQIFINNEHDAVSRKTFPTVNPSTGEVICQVAGSKEDV 60
 Db 1 MSSAPQAPVAPAPLANLKIQTHTKIFINNEHDSVSGKKEPVLNATEEVICHVEEGKADV 60
 QY 61 DKAREGRPAFQSGPWRMDASHGRLNRLADLIERDRTYLAALETLNGKPYVISYL 120
 Db 61 DKAVKAAPQAFQSGPWRMDASERGLNKLADLIERDRTYLAALETLNGKPYVISYL 120
 QY 121 VLDLWVLCRYAGWADKYHGKTIPIGDGFFSTHREPVGVCQGIIPWNEPFLLMQAWKL 180
 Db 121 SPAGGSIKALYAGWADKYHGKTIPIGDGFFSTHREPVGVCQGIIPWNEPFLLMQAWKL 180
 QY 181 GPALATGNVVMKVAEQTPLTALXVANLIKEAGFPFGVNVIPGCGPTAGAAISSHMDID 240
 Db 181 GPALSGNTVVVKPAEQTPLTALHVASLIKEAGFPFGVNVIPGCGPTAGAAISSHMDID 240
 QY 241 KVAFTGSTEIGRIVQVAGSSNLKRVTLLELGKSPNIIIMSDAIMWAVQAHFALFNQ 300
 Db 241 KVAFTGSTEIGRIVQVAGSSNLKRVTLLELGKSPNIIIMSDAIMWAVQAHFALFNQ 300
 QY 301 QCCAGSRTFQEDIEYDFVVRKSVARAKSKVRNPPFQSKTQKSPQVDETFQKILGYINT 360
 Db 301 QCCAGSRTFQEDIEYDFVVRKSVARAKSKVRNPPFQSKTQKSPQVDETFQKILGYINT 360
 QY 361 GKQEGAKLLCGGSIADGNYIFLQPTVFHVGWMTIAKEELFPHVMILKFKETIEEVVGR 420
 Db 361 GKQEGAKLLCGGSIADGNYIFLQPTVFHVGWMTIAKEELFPHVMILKFKETIEEVVGR 420
 QY 421 ANNSTYGLAAAVFTKDLKANYLSQALQAGTVMWNCYDVFCAQSPFGGKMSGSGRELGE 480
 Db 421 ANNSTYGLAAAVFTKDLKANYLSQALQAGTVMWNCYDVFCAQSPFGGKMSGSGRELGE 480


```

Db      361 GKKGAKLECGGGWGNKGFVQVTSNVTDEMIAKEELFGVQVQIMKFKSLDDVIKR 420
      421 ANNSTYGLAAAVFTKDDKANYLSQALQAGTWWVNCYDVFGAQSPPGGYKMSGSGRELGE 480
      422 ANNTYGLAAGVFTKDDRAITVSSALQAGVWVWVNCYMILSAQCFFGGFKMSGNGRELGE 480
QY      481 YGLOAYTEVKTVVKVPQKN 500
      481 HGLYETTELKIVAMKISQKN 500
Db
RESULT 14
S02302
aldehyde dehydrogenase (NAD) (EC 1.2.1.3) 1, cytosolic - horse
C:Species: Equus caballus (domestic horse)
C:Date: 01-Dec-1989 #sequence_revision 31-Jan-1997 #text_change 03-Jun-2002
C:Accession: S02302
R:Von Bahr-Lindstrom, H.; Hempel, J.; Joernvall, H.
Eur. J. Biochem. 141, 37-42, 1984
A:Title: The cytoplasmic isoenzyme of horse liver aldehyde dehydrogenase. Relationship t
A:Reference number: S02302; MUID:84208025; PMID:6723662
A:Accession: S02302
A:Molecule type: protein
A:Residues: 1-500 :VQN-
A:Note: residues in the regions 1-16 and 443-475 were positioned by homology with the hu
C:Complex: homotetramer
C:Function:
A:Description: catalyzes the oxidation of an aldehyde to an acid using NAD+ and water
A:Note: enzymes with this activity are involved in diverse metabolic pathways in various
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
C:Keywords: acetylated amino end, alcohol metabolism; cytosol; homotetramer; liver; NAD;
F:58-322/Domain: aldehyde dehydrogenase homology <ALDH>
F:1/Modified site: blocked amino end (Ser) (probably acetylated) #status experimental
F:268,302/Active site: Glu, Cys #status predicted
F:455/Binding site: NAD (Cys) #status predicted
Query Match      66.1%; Score 1736; DB 1; Length 500;
Best Local Similarity 66.1%; Pred. No. 3.3e-126;
Matches 330; Conservative 65; Mismatches 104; Indels 0; Gaps 0;
QY      2 SAAATQAPAPNQOPEVFCNCFINNEWHDAVSRKTFPTVNPSTGEVICOVAGDKEDVD 61
      1 SSSGTPDPLVLTIDTKQYTKIFINNEWHDSVSGKKFPVPNPAIEKLECEVEGDKEDYN 60
QY      62 KAREGRGAFQGLSPWRMDASHSGRLINRLADLIERDRTYLALETDLONGKPYVTSYLV 121
      61 KAVAAARQAFQIGSPWPTMDASERGLLYKLADLIVERDRLILATMESMNGGKLFSSNAYLM 120
QY      122 DLDWVLKLRYYAGWADKYHGKTTIPDGDFSYTRHEPVGVCQIIPWNPFLLMQAWKLG 181
      121 DLGGGLKTLRYCAGWARKIQGPTIPSDGNFTYTRHEPVGVCQIIPWNPFLLMFLWKTA 180
QY      182 PALATGNVVMKVAEQTPLTALYVANLIKEAGPPGVVNIIVPGFGPTAGAAIASHEDVDK 241
      181 PALSQNTVVVVKPAEQTPLSNLHVATLIKEAGPPGVVNIIVPGFGPTAGAAISSHMDIDK 240
QY      242 VAFSTGTEIGRVIQVAAAGSSNLKRVTLLEGGKSPNIIIMSDADMVAEQAHFALFPNQO 301
      241 VAFSTGTEVGLIKEAAGKSNLKRVTLLEGGKSPFFIVFADADLETALEVTHQALFVHQO 300
QY      302 CCCAGSRFTVQEDYIDFVVRVARAKSRVVGVPFUSKTEQGVDETQFKKILGYNTG 361
      301 CCVAASRTFVPESTYIDFVVRVARAKSRVVGVPFUSKTEQGVDETQFKKILGYNTG 360
QY      362 KOBGAKLQGGJAAADRGYFTQPTVFGVDQDGMTIAKEEIFGPMVMOILFKFTIEEVRG 421
      361 KKEGAKLECGGSPWNGKGYFTQPTVFSNVSDMKIAKEELFGVQVQIMKFKSLDDVIKR 420
QY      422 NNSTYGLAAAVFTKDDKANYLSQALQAGTWWVNCYDVFGAQSPPGGYKMSGSGRELGEY 481
      421 NNTYGLFAGSFTKDDKAITVSSAALQAGTWWVNCYGVVSAQCFFGGFKMSGNGREMGEY 480

```

```

QY      482 GLOAYTEVKTVVKVPQKN 500
      481 GFHYETEVKTVTKISQKN 499
Db
RESULT 15
A55684
aldehyde dehydrogenase (NAD) (EC 1.2.1.3) 6 precursor, salivary - human
C:Species: Homo sapiens (man)
C:Date: 03-Mar-1995 #sequence_revision 31-Jan-1997 #text_change 03-Jun-2002
C:Accession: A55684
R:Hsu, L C; Chang, W C.; Hiraoka, L.; Hsieh, C I.
Genomics 24, 333-341, 1994
A:Title: Molecular cloning, genomic organization, and chromosomal localization of an
A:Reference number: A55684; MUID:95213025; PMID:7698756
A:Accession: A55684
A:Molecule type: mRNA; DNA
A:Residues: 1-512 <HSU>
A:Cross-references: GB:U07919; NID:g995897; PIDN:AAA79036.1; PID:g544482
C:Comment: This isozyme is found at highest levels in saliva, stomach, and kidney and
C:Genetics:
A:Gene: GDB:ALDH6
A:Cross-references: GDB:364103; OMIM:600463
A:Map position: 15q26-15q26
A:Initiator: 107; 108/3; 115/3; 159/1; 179/3; 222/3; 245/3; 411/3; 434/3;
C:Function:
A:Description: catalyzes oxidation of an aldehyde to an acid using NAD+ and water
A:Note: enzymes with this activity are involved in diverse metabolic pathways in vari
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
C:Keywords: NAD; oxidoreductase; saliva
F:780,334/Domain: aldehyde dehydrogenase homology <ALDH>
F:280,314/Active site: Glu, Cys #status predicted
F:467/Binding site: NAD (Cys) #status predicted
Query Match      65.4%; Score 1718; DB 1; Length 512;
Best Local Similarity 65.5%; Pred. No. 8.4e-125;
Matches 323; Conservative 66; Mismatches 104; Indels 0; Gaps 0;
QY      8 AVAPNQOPEVFCNCFINNEWHDAVSRKTFPTVNPSTGEVICOVAGDKEDVDKAREGR 67
      19 ALPRINLELVKFTKIFINNEWHESKSGKFAFCNSTREOICEVEGDKPDVDKAVEAA 78
QY      68 PGAKQLSRPWPMPASHSGPLINPLADLIEPAPTYLALETDLONGKPYVTSYLVLDWVL 127
      79 QVAFQSGSPWPLDALSGPLIHQIADIVRFPATIAALFTMDYGRPFTHAFFIDLEGGCI 138
QY      128 KCLRYAGWADKYHGKTTIPDGDFSYTRHEPVGVCQIIPWNPFLLMQAWKLGSPALATG 187
      139 RTLRYAGWADKIQGTPTIPDNNVVCFTREPIGVCGAITPWNFPDLMVLVKLAPALCCG 198
QY      188 NVVYMKVAEQTPLTALYVANLIKEAGPPGVVNIIVPGFGPTAGAAIASHEDVDKVAFTGS 247
      199 NTWVLKPAEQTPLTALYVLSLIKEAGPPGVVNIIVPGFGPTVCAALSSHPQINKIAFTGS 258
QY      248 TEIGRVIOVAAAGSSNLKRVTLLEGGKSPNIIIMSDADMVAEQAHFALFPNQOCCAGS 307
      259 TEVGLKVEAKSNLKKVTLLEGGKSPCLVCAALDLDLAVECAHGVFENQGCCTAAS 318
QY      308 KTFVQRPIDYDFVVRVARAKSRVVGVPFUSKTEQGVDETQFKKILGYNTGKUEGAK 367
      319 RVFVEQVYSEFVVRVARAKSRVVGVPFUSKTEQGVDETQFKKILGYNTGKUEGAK 378
QY      368 LLCGGJAAADRGYFTQPTVFGVDQDGMTIAKEEIFGPMVMOILFKFTIEEVRGANNSTYG 427
      379 LECGGSAMEDKGLFKPTVSEVTDNMRINAKKEIFGVPQIPILFKFSIEEVIKRNSTDYG 438
QY      428 LAAAVFTKDDKANYLSQALQAGTWWVNCYDVFGAQSPPGGYKMSGSGRELGEYGLQAYT 487
      439 LTAAVFTKDDKALKLASALESGTVMVINCYNALYAQAEPFGGFKMSGNGRELGEYALABYT 498
QY      488 EVKTVTVKVPQKN 500
      499 EVKTVTVKLGDKN 511

```

Search completed: June 24, 2003, 10:18:30
Job time : 20.1339 secs

GenCore version 5.1.6

Copyright (c) 1993-2003 Compugen Ltd

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:02:55; Search time 10.0735 seconds

(without alignments)

2059.300 Million cell updates/sec

Title: US-09-830-751-4

Perfect score: 2625

Sequence: 1 MSAATGAVPAPNQPEVF YGLQAYIEVKIVTKVPQKN 500

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY

Result No.	Score	Query Match	Length	DB ID	Description
1	2572	98.0	517	1	DHAM_HUMAN
2	2563	95.4	500	1	DHAM_MOUSE
3	2500	95.2	519	1	DHAM_RAT
4	2480	94.5	519	1	DHAM_MOUSE
5	2436	92.8	520	1	DHAM_BOVIN
6	2434	92.7	500	1	DHAM_MOUSE
7	1929	73.5	517	1	DHAM_HUMAN
8	1802	68.6	509	1	DHAM_CHICK
9	1795	68.4	499	1	DHAM_RAT
10	1794	68.3	499	1	DHAM_MOUSE
11	1790	68.2	499	1	DHAM_HUMAN
12	1771	67.5	499	1	DHAM_CHICK
13	1769	67.4	500	1	DHAM_HUMAN
14	1758	67.0	500	1	DHAM_SHEEP
15	1752	66.7	500	1	DHAM_MOUSE
16	1752	66.7	500	1	DHAM_RAT
17	1751	66.7	500	1	DHAM_BOVIN
18	1736	66.1	500	1	DHAM_MOUSE
19	1718	65.4	512	1	DHAM_HUMAN
20	1711	65.2	500	1	DHAM_RAT
21	1706	65.0	501	1	DHAM_MOUSE
22	1697	64.6	501	1	DHAM_MOUSE
23	1587	60.5	497	1	DHAM_MOUSE
24	1526	58.2	494	1	DHAM_MOUSE
25	1499	57.1	495	1	DHAM_MOUSE
26	1425.5	54.3	497	1	DHAM_MOUSE
27	1422.5	54.2	511	1	DHAM_MOUSE
28	1353.5	51.6	497	1	DHAM_MOUSE
29	1305.5	49.7	496	1	DHAM_MOUSE
30	1269	48.3	495	1	DHAM_MOUSE
31	1258	47.9	500	1	DHAM_MOUSE
32	1235.5	47.1	519	1	DHAM_MOUSE
33	1199.5	45.7	498	1	DHAM_MOUSE

RESULT 1

DHAM_HUMAN

ID	DHAM_HUMAN	STANDARD	PRT	517 AA
AC	P05091	Q03639		
DT	13-AUG-1987	(Rel. 05, Created)		
DT	01-JAN-1990	(Rel. 13, Last sequence update)		
ET	15-JUN-2002	(Rel. 41, Last annotation update)		
EF	Aldehyde dehydrogenase, mitochondrial precursor (EC 1.2.1.3) (ALDH)			
DE	class 2) (ALDH1) (ALDH-E2).			
GN	ALDH2 OR ALDH			
OS	Homo sapiens (Human).			
OC	Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi;			
OC	Mammalia, Eutheria, Primates, Catarrhini, Hominoidea, Homo.			
OX	NCBI_TaxID=9606;			
PN	[1]			
PP	SEQUENCE FROM N.A.			
RX	MEDLINE=88256152; PubMed=3838413;			
RA	Hsu L.C., Bendel R.E., Yoshida A.;			
RT	"Genomic structure of the human mitochondrial aldehyde dehydrogenase			
RT	gene.";			
RL	Genomics 2:57-65(1988).			
RN	[2]			
RC	SEQUENCE OF 1A-517			
TX	TISSUE=Liver;			
TX	MEDLINE=86055846; PubMed=4061146;			
RT	Hempel J., Kaiser R., Joernvall H.;			
RT	"Mitochondrial aldehyde dehydrogenase from human liver. Primary			
RT	structure, differences in relation to the cytosolic enzyme, and			
RT	functional correlations.";			
PL	Eur. J. Biochem. 153:13-28(1985).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Muscle;			
EX	MEDLINE=87174836; PubMed=3562250;			
RA	Braun T., Rober E., Singh S., Agarwal D.P., Goedde H.W.;			
PT	"Isolation and sequence analysis of a full length cDNA clone coding			
PT	for human mitochondrial aldehyde dehydrogenase.";			
PL	Nucleic Acids Res. 15:3179-3179(1987).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Muscle;			
EX	MEDLINE=87219991; PubMed=3582651;			
RA	Braun T., Rober E., Singh S., Agarwal D.P., Goedde H.W.;			
RT	"Evidence for a signal peptide at the amino terminal end of human			
RT	mitochondrial aldehyde dehydrogenase.";			
PL	FEBS Lett. 215:233-236(1987).			
RN	[5]			
PP	REVISIONS TO N-TERMINUS			
RA	Braun T., Rober E., Singh S., Agarwal D.P., Goedde H.W.;			
PL	FEBS Lett. 233:440-440(1988).			
RN	[6]			
PP	SEQUENCE FROM N.A.			
RC	TISSUE=Lymph;			
RA	Straussberg R.;			
RT	Submitted (FEF-2001) to the EMBL/GenBank/DBJ databases.			
RN	[7]			

P40047 saccharomyc
P75891 homo sapien
P28937 rattus norv
P54115 saccharomyc
P47771 saccharomyc
P54114 saccharomyc
P71016 bacillus su
P28237 beta vulgar
P17202 spinacia ol
P42757 atiplex ho
P17445 escherichia
P24883 escherichia

```

or send an email to license@usf-sid.ch).
CC
DR EMBL: M20456; AAA51693.1; JOINED.
DR EMBL: M20445; AAA51693.1; JOINED.
DR EMBL: M20445; AAA51693.1; JOINED.
DR EMBL: M20446; AAA51693.1; JOINED.
DR EMBL: M20447; AAA51693.1; JOINED.
DR EMBL: M20448; AAA51693.1; JOINED.
DR EMBL: M20449; AAA51693.1; JOINED.
DR EMBL: M20450; AAA51693.1; JOINED.
DR EMBL: M20451; AAA51693.1; JOINED.
DR EMBL: M20452; AAA51693.1; JOINED.
DR EMBL: M20453; AAA51693.1; JOINED.
DR EMBL: M20454; AAA51693.1; JOINED.
DR EMBL: X05403; CAA28900.1; -.
DR EMBL: Y00109; CAA68290.1; -.
DR EMBL: BC002967; AAH02967.1; -.
DR EMBL: K03001; AAB59500.1; -.
DR EMBL: M26760; AAA51694.1; -.
DR EMBL: M54931; AAA62825.1; ALT_FRAME.
DR PIR: S00804; DEH0E2.
DR PIR: A29975; A29975.
DR PDB: 1CW3; 1C-WAN-00.
DR Genew: HGNC:404; ALDH2.
DR MIM: 100650; -.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldedh.1
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS.1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU.1.
DR OXIDOREDUCTASE, NAD; Mitochondrion; Transist peptido; Polymorphism;
KW 3D-structure.
KW TRANSIT 1 17
KW CHAIN 18 517
KW NP_BIND 262 267
KW ACT_SITE 285 285
KW ACT_SITE 319 319
KW VARIANT 337 337
KW VARIANT 496 496
KW VARIANT 504 504
KW VARIANT 504 504
KW CONFLICT 7 12
KW CONFLICT 7 26
KW CONFLICT 18 18
KW CONFLICT 80 85
KW CONFLICT 216 216
KW CONFLICT 218 218
KW CONFLICT 247 247
KW CONFLICT 380 380
KW CONFLICT 462 463
KW SEQUENCE 517 AA; 56381 MW; F8F74D40285A00E CRC64;
SEQUENCE 517;

```

Query Match	98.0%;	Score	No. 476-140;	Indels	Gaps
Best Local Similarity	98.2%;	Pred.	No. 476-140;		
Matches	491;	Conservative	Mismatches	7;	
1	MSAAATQAVAPNQOPEVFCNQIFNNKSHVAVSRKTFPTVNFSTGEVTCQVAEGCKEDW	60			
17	LSAAATQAVAPNQOPEVFCNQIFNNKSHVAVSRKTFPTVNFSTGEVTCQVAEGCKEDW	76			
61	DKAREGRGAPQLGSGPWRKMDASHGRGLLNRLALDERKTYLAALFTLLNKPVVISYL	120			
77	DKAVKAAAPQLGSGPWRKMDASHGRGLLNRLALDERKTYLAALFTLLNKPVVISYL	146			
121	VLDLWVLCLELYAGWADKYHGKTIPTDGLDFESYTRHEFPVGVCGQTPNWFLLMLAKKL	180			
137	VLDLWVLCLELYAGWADKYHGKTIPTDGLDFESYTRHEFPVGVCGQTPNWFLLMLAKKL	196			
181	GPALATGNVVMKVAEQITFIALKXVNLKFCNCFPPGVNIVPCEKPTAGAAATASHEDWD	240			

Query Match

Best Local S

Matches 491

—

17

61

77

127

121

137

107

1917

[REDACTED]

SEQUENCE OF 119-517 FROM N A

TISSUE=Liver;

MEDLINE=85216574; PubMed=2987944;

RA Hsu L.-C., Tani K., Fuli-yoshi T., Kurachi K., Yoshida A.;

RT "Cloning of cDNAs for human aldehyde dehydrogenases 1 and 2.;"

proc. Natl. Acad. Sci. U.S.A. 82:3771-3775(1985).

RA [8]

SEQUENCE OF 119-517 FROM N A

RA MEDLINE=85252089; PubMed=4015823;

RA Yoshida A., Ikawa M., Hsu L.-C., Tani K.;

RT "Molecular abnormality and cDNA cloning of human aldehyde

RT dehydrogenases.;"

RA Alcohol 2:103-106(1985).

RA [9]

SEQUENCE OF 214-500 FROM N A.

TISSUE=Liver;

RA MEDLINE=87279033; PubMed=3610592;

RA Agarwal D.P., Goedde H.W.;

PT "Human aldehyde dehydrogenase isozymes and alcohol sensitivity.;"

RA Isozymes Curr. Top. Biol. Med. Res. 16:21-48(1987).

RA [10]

DESCRIPTION OF ORIGIN OF CONFLICTS BETWEEN REF.2 AND DNA SEQUENCES.

RA MEDLINE=88005159; PubMed=3653404;

RA Hempel J., Hoeseg J.-O., Joernvall H.;

RA "Mitochondrial aldehyde dehydrogenase. Homology of putative targeting

RT sequence to that of carbamyl phosphate synthetase I revealed by

RT correlation of cDNA and protein data.;"

RA FEBS Lett. 222:95-98(1987).

RA [11]

VARIANT LYS-504.

RA MEDLINE=84119449; PubMed=6582480;

RA Yoshida A., Huang I.-Y., Ikawa M.;

RT "Molecular abnormality of an inactive aldehyde dehydrogenase variant

RT commonly found in Orientals.;"

Proc. Natl. Acad. Sci. U.S.A. 81:258-261(1984).

RA [12]

VARIANT LYS-496.

RA MEDLINE=96119362; PubMed=8561277;

RA Novorodovsky A., Tsai S.-J., Goldfarb L., Peterson P., Long J.C.,

RA Goldman D.;

RT "Mitochondrial aldehyde dehydrogenase polymorphism in Asian and

RT American Indian populations: detection of new ALDH2 alleles.;"

RA Alcohol. Clin. Exp. Res 19:1105-1110(1995).

RA [13]

X-RAY CRYSTALLOGRAPHY (2.58 ANGSTROMS).

RA MEDLINE=20095857; PubMed=10631996;

RA Ni L., Zhou J., Hurley T.D., Weiner H.;

RT "Human liver mitochondrial aldehyde dehydrogenase: three-dimensional

RT structure and the restoration of solubility and activity of chimeric

RT forms.;"

RA Protein Sci. 8:2784-2790(1999).

CC -1- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.

CC -1- PATHWAY: Ethanol utilization; second step.

CC -1- SUBUNIT: HOMOTETRAMER.

CC -1- DISEASE: A VERY HIGH INCIDENCE OF ACUTE ALCOHOL INTOXICATION IN

CC ORIENTALS AND SOUTH AMERICAN INDIANS, AS COMPARED TO CAUCASIANS,

CC CAN BE DIRECTLY ATTRIBUTED TO AN ENZYMATICALLY IMPAIRED ALDH

CC ISOZYME, THE INACTIVE VARIANT (ALLELE 2 OR ALDH2*2) IS DUE TO

CC SINGLE AMINO ACID EXCHANGE.

CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.

CC -1- CAUTION: THE CONFLICTS BETWEEN THE SEQUENCE DETERMINED IN REF.1

CC AND REF.2 AND THAT IN REF.3 AND 4 ARE PROBABLY ALL DUE TO

CC FRAMESHIFT OR SEQUENCING ERRORS AS DESCRIBED IN REF.5 AND REF.6;

CC THE SEQUENCE DESCRIBED IN REF.9 DIFFERS FROM THAT SHOWN DUE TO TWO

CC FRAMESHIFTS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

```

Db 197 GPALATGNVVMKVAEOTPLTALYVANLIKEAGFPVGVNIVPGFGP/TAGAAIASHEDVD 256
QY 241 KVAFTGSTEIGRVTOVAGSSNLKRVTLLEGGSPNIIIMSDADMDWAVEQAHFALFFNOG 300
Db 257 KVAFTGSTEIGRVTOVAGSSNLKRVTLLEGGSPNIIIMSDADMDWAVEQAHFALFFNOG 316
QY 301 QCCAGSRTFVQEDIDYDFVVRVARAKSRVGVNPFDSKTEQGPQVDETOFKKILGY INT 360
Db 317 QCCAGSRTFVQEDIDYDFVVRVARAKSRVGVNPFDSKTEQGPQVDETOFKKILGY INT 376
QY 361 GKQEGAKLLCGGIIAARGYFIQPTVFGDVJUDKMIIAKEEIPFGVMJLKFKETIEBVVGR 420
Db 377 GKQEGAKLLCGGIIAARGYFIQPTVFGDVJUDKMIIAKEEIPFGVMJLKFKETIEBVVGR 436
QY 421 ANNSTYGLAAAVFTKDLKANYLSOALQAGTVMVNCYDVFGCAQSPGGYKMSGSGRELGE 480
Db 437 ANNSTYGLAAAVFTKDLKANYLSOALQAGTVMVNCYDVFGCAQSPGGYKMSGSGRELGE 496
QY 481 YGLQAYTEVKTIVTKVPQKN 500
Db 497 YGLQAYTEVKTIVTKVPQKN 516

RESULT 2
DHAM_MESAU
ID DHAM_MESAU STANDARD: PPT: 500 AA
AC P81178:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Aldehyde dehydrogenase, mitochondrial (EC 1.2.1.3) (ALDH class 2)
DE (ALDH1) (ALDH-E2).
GN ALDH2.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE.
RC TISSUE=Liver;
PX MEDLINE=98034175; PubMed=9369242;
RA Hjeltnist L., Lundgren R., Norin A., Joernvall H., Vallee B.,
RA Kiyosaki A., Keung W.M.;
ET "Class 2 aldehyde dehydrogenase. Characterization of the hamster
ET enzyme, sensitive to dieldrin and conserved within the family of
ET multiple forms."
RL FEBS Lett. 416:99-102(1997)
CC -|- CATALYTIC ACTIVITY: An aldehyde + NAD(+) -> H(2)O + an acid + NADH.
CC -|- PATHWAY: Ethanol utilization; second step.
CC -|- SUBUNIT: HOMOTETRAMER
CC -|- SUBCELLULAR LOCATION: Mitochondrial matrix
CC -|- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
DR HSP: P05091; ICB3
DR InterPro: Ipp002086; Aldehyde_dehyd.
DR Pfam: PF00171; aldh; 1
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_G1; 1
DR PROSITE: PS00670; ALDEHYDE_DEHYDR_GYS; 1
KW Oxidoreductase; NAD; Mitochondrion.
FT ACT_SITE 245 250 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 268 268 BY SIMILARITY.
FT ACT_SITE 302 302 BY SIMILARITY.
SQ SEQUENCE 500 AA: 54334 MW; 8881886AB04F493F CRC64;

Query Match 95.4%; Score 2503; DB 1; Length 500;
Best local Similarity 95.0%; Pred. No. 1,6e-184;
Matches 474; Conservative 11; Mismatches 14; Indels 0; Gaps 0;

QY 2 SAAATGAPAPNQGFEVFTNIFINNWHIAVSKHFFVNPSTGEVLCQVABEDKEDVD 61
Db 1 SAAATGAPAPNQGFEVFTNIFINNWHIAVSKHFFVNPSTGEVLCQVABEDKEDVD 60
QY 62 KAPFTGTCATGSGFWPMTWASHSTPLINPLAFLIEPQYLAALETLLNGKPYVSYLV 121

```

```

Db 51 KAVFAARAAAFQSGSPWPMASDECRLLNPLADLJERDRTYLAALETLLNGKPYVSYLV 120
QY 122 DLDMY: KCLPYVAAWAKYBKRTIPIDGDFEFTPTPEFVGVQSCIIIPWNPFLIMQAWPLG 181
Db 121 DLDMY: KCLRYVAGWADKYHGKTIPIIDGDFEFTPTPEFVGVQSCIIIPWNPFLIMQAWPLG 180
QY 182 PALATGNVVMKVAEOTPLTALYVANLIKEAGFPVGVNIVPGFGP/TAGAAIASHEDVDK 241
Db 181 PALATGNVVMKVAEOTPLTALYVANLIKEAGFPVGVNIVPGFGP/TAGAAIASHEDVDK 240
QY 242 VAFVGTSTIEIGSVLVAAASSSNLKPVTLLEGGSPNIIIMSDADMDWAVEQAHFALFFNOG 301
Db 241 VAFVGTSTIEIGSVLVAAASSSNLKPVTLLEGGSPNIIIMSDADMDWAVEQAHFALFFNOG 300
QY 302 CCCAGSRTFVQEDIDYDFVVRVARAKSRVGVNPFDSKTEQGPQVDETOFKKILGY INT 361
Db 301 CCCAGSRTFVQEDIDYDFVVRVARAKSRVGVNPFDSKTEQGPQVDETOFKKILGY INT 360
QY 362 KQCAKTIICGGGIIAARGYFIQPTVFGDVJUDKMIIAKEEIPFGVMJLKFKETIEBVVGR 421
Db 361 KQCAKTIICGGGIIAARGYFIQPTVFGDVJUDKMIIAKEEIPFGVMJLKFKETIEBVVGR 420
QY 422 NNSTYGLAAAVFTKDLKANYLSOALQAGTVMVNCYDVFGCAQSPGGYKMSGSGRELGEY 481
Db 421 NNSTYGLAAAVFTKDLKANYLSOALQAGTVMVNCYDVFGCAQSPGGYKMSGSGRELGEY 480
QY 482 GLQAYTEVKTIVTKVPQKN 500
Db 481 GLQAYTEVKTIVTKVPQKN 499

RESULT 3
DHAM_PAT
ID DHAM_PAT STANDARD: PPT: 519 AA.
AC P11884:
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15 JUN 2002 (Rel. 41, Last annotation update)
DE Aldehyde dehydrogenase, mitochondrial precursor (EC 1.2.1.3) (ALDH
DE class 2) (ALDH1) (ALDH-E2).
GN ALDH2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RC MEDLINE=89210865; PubMed=2540003;
RA Farres J., Guan K.-L., Weiner H.;
RA Farres J., Guan K.-L., Weiner H.;
ET "Primary structures of rat and bovine liver mitochondrial aldehyde
ET dehydrogenases deduced from cDNA sequences."
RL Eur. J. Biochem. 180:57-74(1989).
RN [2]
RP SEQUENCE OF 1-29 FROM N.A.
RC TISSUE=Liver;
RC MEDLINE=89144717; PubMed=3442060;
RA Farres J., Guan K.-L., Weiner H.;
RA Farres J., Guan K.-L., Weiner H.;
ET "Sequence of the signal peptide for rat liver mitochondrial aldehyde
ET dehydrogenase."
RL Biochem. Biophys. Res Commun 150:1083-1087(1988).
RN [3]
RP SEQUENCE OF 1-19.
RC TISSUE=Liver;
RC MEDLINE=91378548; PubMed=1898068;
RA Jeng J., Weiner H.;
RA Jeng J., Weiner H.;
ET "Purification and characterization of catalytically active precursor
ET of rat liver mitochondrial aldehyde dehydrogenase expressed in
ET Escherichia coli."
RL Arch. Biochem. Biophys. 289:214-222(1991).
CC -|- CATALYTIC ACTIVITY: An aldehyde + NAD(+) -> an acid + NADH.
CC -|- PATHWAY: Ethanol utilization; second step.

```



```

CC (BY SIMILARITY).
CC -|- CATALYTIC ACTIVITY: An aldehyde + NAD(+) -> H(2)O + an acid + NADH
CC -|- PATHWAY: Ethanol utilization, second step.
CC -|- SUBUNIT: HEMETPAP.
CC -|- SUBCELLULAR LOCATION: Cytoplasmic.
CC -|- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M31984; AA51692.1; JOINED
DR EMBL: M31982; AA51692.1; JOINED
DR EMBL: M31983; AA51692.1; JOINED
DR EMBL: M31984; AA51692.1; JOINED
DR EMBL: M31985; AA51692.1; JOINED
DR EMBL: M31986; AA51692.1; JOINED
DR EMBL: M31987; AA51692.1; JOINED
DR EMBL: M31988; AA51692.1; JOINED
DR EMBL: M31989; AA51692.1; JOINED
DR EMBL: M31990; AA51692.1; JOINED
DR EMBL: M31991; AA51692.1; JOINED
DR EMBL: M31992; AA51692.1; JOINED
DR EMBL: AF003341; AAC51652.1;
DR EMBL: K03000; AA51695.1;
DR EMBL: M26761; AA35519.1;
DR EMBL: M561345; AAD13925.1;
DR PIR: A33371; DEH01.
DR HSSP: P51977; 1BX5.
DR SWISS-2DPAGE: P00352; HUMAN.
DR Genew: HGNC:402; ALDH1A1.
DR MIM: 100640;
DR InterPro: IPR00286; Aldehyde_d-hydr
DR Pfam: PF00171; aldehyd; 1.
DR PROSITE: PS00670; ALDEHYDE_DEHYDR_GYS; 1
DR PROSITE: PS00667; ALDEHYDE_DEHYDR_GLIF; 1.
KW oxidoreductase; NAD; Acetylation
FT INIT_MET 0
FT MOD_RES 1
FT NP_BIND 245 250 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 248 268 POTENTIAL.
FT ACT_SITE 302 302 POTENTIAL.
FT CONFLICT 120 120 N -> S (IN REF. 2).
FT CONFLICT 161 161 V -> I (IN REF. 3 AND 4).
SQ SEQUENCE 500 AA; 54730 MW; 82645414564749677 Cys64;
Query Match 67.4%; Score 1769; DB 1; Length 500;
Best Local Similarity 66.9%; Pred. No. 3.3e-128;
Matches 334; Conservative 67; Mismatches 98; Indels 0; Gaps 0.
QY 2 SAAATQAVFAINQOIEVEFCNO;PINNEWHUAVSKKFEIVNLSIGEVLTAVABGJKEWQ;61
Db 1 SSSGTPLPVLTLTKYTKIPINNEWHUAVSKKPPVFNPAIEBGLQVEGKEDVD 60
QY 62 KARETRKPAQAGSPPWRMIAASHKAPLLNPLAMIEKQFTYLALETLENKPYVLSLV 121
Db 61 KAVKAAQCAFGTSPWPTMPASEPGPIIYKIDIEPDLIATMESMNGEKLISNAYLN 120
QY 122 DLDWVLCFLRYACWAKYKTKTPTDGFSTFSTHEFVGVQQTITWNPFLIMQAWKLG 181
Db 121 DLDWVLCFLRYACWAKYKTKTPTDGFSTFSTHEFVGVQQTITWNPFLIMQAWKLG 180
QY 182 PALATGNVVMKVAEOTPLTALYVANLIKEAGPPGVNVIYPCGPTAGAAIASHRDVDK 241
Db 181 PALSCGNVVMVVKPAQVPTPLTALHVASLIKAGPPGVNVIYPCGPTAGAAISSHMLDK 240
QY 242 VAFSTSTELISPVIGAAVSSNKKVTLFEQSKSFNIIIMSDAMWAVESAHLFALFNQQQ 301
Db 241 VAFSTSTFVSKLKEAAQCKSNIAKPVTLFEQSKSNIVIALADLIINAVFAHHGVFVHQDQ 300

```

```

QY 402 QVAGSKFTFQDEIYDFEFVVSAPAKSPVVNPHLSKFTGSPQVLEFQKKLLGYINIG 361
Db 401 QVIAASRIFVEESYDFEFVVSAPAKKYIDNPHIIPSTVQAPQPIIKEDYKLLDILISG 360
QY 362 KQECARLEGGGGIAADGGYFQAPVFCWQJPMTEIAKEETPGIVMAJLKFRTIFFVVGKA 421
Db 361 KKECAKLEGGGGWGNKGYFVQPVFSNVTHMPJAKEELPQEVVLTIMKFSJLQVTKPA 420
QY 422 NNSTYGLAAAVFTKCIKANKYLSQALQAGIVWVWYVQVFAISPFQSYKMSGSPFLRY 481
Db 421 NNIFYGLSAGVFKDIQKALLISSALQAGIVWVWYVWSAQCPFGGFKMSGRELGEY 480
QY 482 GLQAVTEVKTIVTKVPKN 500
Db 481 GFHEYTEVKTIVTKISOKN 499
RESULT 14
DHAL_SHEEP
ID DHAL_SHEEP STANDARD; PRT; 500 AA.
AC P51977;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Aldehyde dehydrogenase 1A1 (F1.1.2.1.3) (Aldehyde dehydrogenase,
DE cytosolic) (ALDH class 1) (ALDH1) (ALDH-E1).
GN ALDH1A1 OP ALDH1.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Fuminiatia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
FX MEDLINE=96062935; PubMed=7484149;
RA Stayner C.K., Tweedie J.W.;
RT "Cloning and characterisation of the cDNA for sheep liver cytosolic
RT aldehyde dehydrogenase."
PL Adv. Exp. Med. Biol. 372:61-66(1995).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.35 ANGSTROMS).
RC TISSUE=Liver;
FX MEDLINE=99081884; PubMed=9852807;
RA Moore S.A., Baker H.M., Rlythe T.J., Kitson K.E., Kitson T.M.,
RA Baker E.N.;
RT "Sheep liver cytosolic aldehyde dehydrogenase: the structure reveals
RT the basis for the retinal specificity of class 1 aldehyde
RT dehydrogenases."
RL Structure 6:1541-1551(1998).
CC -|- FUNCTION: BINDS FIVE RETINAL AND CELLULAR RETINOL-MININ; PROTEIN-
CC BIND RETINAL. CAN CONVERT XENOBOTIC RETINALDEHYDE TO RETINOIC ACID.
CC -|- CATALYTIC ACTIVITY: An aldehyde + NAD(+) -> an acid + NADH.
CC -|- PATHWAY: Ethanol utilization, second step.
CC -|- SUBUNIT: HEMETPAP.
CC -|- SUBCELLULAR LOCATION: Cytoplasmic.
CC -|- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U12761; AAA85435.1;
DR PDB: 1UXS; 27-APP-99.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldehyd; 1.
DR PROSITE: PS00670; ALDEHYDE_DEHYDR_GYS; 1.
DR PROSITE: PS00667; ALDEHYDE_DEHYDR_GLIF; 1

```

```

KW Oxidoreductase; NAD; Acetylation; 3D-structure.
FT INIT_MET 0 0 BY SIMILARITY.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT NP_BIND 245 250 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 268 302
FT ACT_SITE 302 302
SQ SEQUENCE 500 AA; 54693 MW; 68B8971968CEB8F8 CRC64;

Query Match 67.0%; Score 1758; DB 1; Length 500;
Best Local Similarity 66.1%; Pred. No. 2,3e-127;
Matches 332; Conservative 67; Mismatches 100; Indels 0; Gaps 0;

QY 2 SAATAOAVPAPNOQPEVFCNQIFINNEHDAVSRRKIFPTVNPSTGCEVICQVAEGDKEDVD 61
DB 1 SSSAMPDVPAPLITNIQFKYKIFINNEHSSVSGKKFPVNPATEEKLCEVEBEGDKEDVD 60
QY 62 KAREGRGAFQIGSPWRMDASHSGRLNRLADLIFDRDTYLALETLDNCKPVVSYLV 121
DB 1 KAVKAARQAFQIGSPWRTMDASRGPFLNKLADLIFDRDLATMEAMNGGKLFENAYLM 120
QY 122 DLDMVLKCLRYVAGMADKYHGKTIPIDGDFPSYTRHEPVGCGOIIIPNFPPLMQAMKLG 181
DB 121 DLGGCIKILYACAWAKLQCKRIIPMGNFFIYKSEPVGCGQIIPNFPPLMFLWKLG 180
QY 182 PALATGNVVMKVAEOTPLTALYVANLIKAEAGFPFGVNVNIVPGSGPTAGAAIASHEDYDK 241
DB 181 FALSCGNTVVKFAEOTPLTALHMSGLIKAEAGFPFGVNVNIVPGSGPTAGAAISSHIMVUK 240
QY 242 VAFGSGTEICRVLOVAGSSNLIKRVTLGLGKSPNIMSDADMWAVEQAHFALFPNGQ 301
DB 241 VAFGSGTEVCKIKREAAKSNIKPVSLGLGKSPCIFVADADLUNAVEFAHQGVFYHQG 300
QY 302 CCAGSKIEVQEDIDYDFVRSVARAKSRVGVGNPDSPKTEQGPQVDFOTQFKKILGYINTG 361
DB 301 CCIAASKIEVRSIYDFVRSVERAKKYLGNPLTPGVSGOQPIDKREYKILDLIESG 360
QY 362 KQEGAKLLGGGIAADKGYFIQPTVFGDVQDGMTIAKEEIRGPVWQILKFKTIEEVVGRA 421
DB 361 KREGAKLEGGGPGWNGKGYFIQPTVFSVDVDMIRIAKEEIRGPVQQIMKFKSLDDVTKRA 420
QY 422 NNSYGLAAAVFTKDLIKANYLSOALQACTVWVNCYDFVFAQSGSPGKMSNKKPLGEY 481
DB 421 NTFYGLSAGIFINDIKAITVSSALQSGTVWVNCYSVVSACQPGFGKMSGNGRELGEY 480
QY 482 GLQAYTEVKTIVKVPQKN 500
DB 481 GPHEYTEVKTIVTKISOKN 499

RESULT 15
DHAI_MOUSE: STANDARD; PRT; 500 AA.
AC P24549;
DT 01-MAR-1992 (Rel. 21, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Aldenhyde dehydrogenase IAI (EC 1.2.1.3) (Aldenhyde dehydrogenase,
DE cytosolic) (ALDH class I) (ALDH1) (ALDH-E1).
CN ALDH1A1 OR ALDH1 OR AHD2 OR AHD-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
PC STPAIN=BALB/c, and C57BL/6; TISSUE=Liver;
FX MEDLINE=91276281; PubMed=2055490;
RA Rongoparut P. Weaver S.;
RT "Isolation and characterization of a cytosolic aldehyde
RT dehydrogenase-encoding cDNA from mouse liver.";
KL Gene 101:261-265(1991).
RN [2]
RP SEQUENCE FROM N.A.

```

```

KC STRAIN=BALB/c, C57BL/6J, and 129/REV; TISSUE=Liver;
KC MEDLINE=95085815; PubMed=7994664;
RA Bond S.L., Singh S.M.;
RT "DNA sequence analysis of the cytosolic acetaldehyde dehydrogenase
RT gene (Ahd-2) in mouse strains with variable ethanol preferences.";
RL Biochem. Med. Metab. Biol. 52:155-159(1994).
CC -1- FUNCTION: IN ADDITION TO THE ACTIVITY ON ACETALDEHYDE AND RELATED
CC SUBSTRATES, IS ALSO INVOLVED IN THE OXIDATION OF ALDEHYDES DERIVED
CC FROM BIOGENIC AMINES SUCH AS EPINEPHRINE AND NOREPINEPHRINE, AS
CC WELL AS THE ALDEHYDES GENERATED VIA LIPID PEROXIDATION. BINDS FREE
CC RETINAL AND CELLULAR RETINOL-BINDING PROTEIN-BIND RETINAL. CAN
CC CONVERT/OXIDIZE RETINALDEHYDE TO RETINOIC ACID (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O -> an acid + NADH.
CC -1- PATHWAY: Ethanol utilization; second step.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE LIVER, LUNG, AND TESTES. IS
CC APPARENTLY NOT EXPRESSED AT DETECTABLE LEVELS IN KIDNEY, STOMACH,
CC OVARY, HEART, AND BRAIN.
CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb.slb.ch/announcements> or send an email to license@isb.slb.ch).

```

EMBL: M74570; AAA37202.1;
EMBL: M74571; AAA37204.1;
EMBL: S75714; AAB32754.2;
EMBL: S77047; NOT_ANNOTATED_CDS.
PIR: J01004; J01004.
ISSP: P51977; IHSX.
DR SWISS-2DPAGE; P24549; MOUSE.
DR MGI; MGI:1453450; Alldhal.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF00171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00087; ALDEHYDE_DEHYDR_GLD; 1.
KW Oxidoreductase; NAD; Acetylation.
FT INIT_MET 0 0 BY SIMILARITY.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT NP_BIND 245 250 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 268 302 BY SIMILARITY.
FT ACT_SITE 302 302 BY SIMILARITY.
FT BINDING 299 304 TO ANTARUSE.
FT CONFLICT 7 7 A -> R (IN REF. 1).
FT CONFLICT 44 44 T -> S (IN REF. 2).
FT CONFLICT 50 50 H -> Q (IN REF. 2).
FT CONFLICT 86 86 R -> C (IN REF. 1).
FT CONFLICT 457 457 I -> M (IN REF. 2).
SQ SEQUENCE 500 AA; 54318 MW; 3E428154E7214B54 CRC64;

```

Query Match 66.7%; Score 1752; DB 1; Length 500;
 Best Local Similarity 66.1%; Pred. No. 6,5e-127;
 Matches 330; Conservative 69; Mismatches 100; Indels 0; Gaps 0;

```

QY 2 SAATAOAVPAPNOQPEVFCNQIFINNEHDAVSRRKIFPTVNPSTGCEVICQVAEGDKEDVD 61
DB 1 SSSAMPDVPAPLADLKIQTKIFINNEHNSVSKKFFVLPATFEVLCIVEEGDKEDVD 60
QY 62 KAREGRGAFQIGSPWRMDASHSGRLNRLADLIFDRDTYLALETLDNCKPVVSYLV 121
DB 61 KAVKAARQAFQIGSPWRTMDASRGPFLNKLADLIFDRDLATMEAMNGKLFENAYLM 120
QY 122 DLDMVLKCLRYVAGMADKYHGKTIPIDGDFPSYTRHEPVGCGOIIIPNFPPLMQAMKLG 181
DB 121 DLGGCIKILYACAWAKLQCKRIIPMGNFFIYKSEPVGCGQIIPNFPPLMFLWKLG 180
QY 182 PALATGNVVMKVAEOTPLTALYVANLIKAEAGFPFGVNVNIVPGSGPTAGAAIASHEDYDK 241
DB 181 FALSCGNTVVKFAEOTPLTALHMSGLIKAEAGFPFGVNVNIVPGSGPTAGAAISSHIMVUK 240

```



```
Db 181 PALSOGNTVVVKPAFQPTLTALHLASLIKEAGFPCCVVNTVPGYGPAGAAISSHMDVK 240
QY 242 VAFGTGSTFGRVITQVAAGSSNLKRVTLLELGKSPNTIMSDADMWAVEGAHFALFNQJ 301
Db 241 VAFGTGSTQVGLIKEAAGKSNLKRVTLELGKSPCIVFADADLDIAVEFAHGVYHQJ 300
QY 302 CTCAGSPFVQEDHYDEFVVPVAPAKSPVVGNEPFGSKTEQGHCVGDEIQFKKILGYINTG 361
Db 301 CCVAASRIFVEESYDEVRKRSVERAKKYVLGNPLTPGINQGPQIDKEQHDKILDLESG 360
QY 362 KQRTAKLIGGSGIAADPGYFIQPIVPGHJGDMTIAKEELFGPVMQILKPKTLEEYVGP 421
Db 361 KKEGAKIEGGRGFWNGGFVQPIVFSNVILEMFIAKEELFGPVMQILKPKSVDDVIKPA 420
QY 422 NNSTVGIAAAVFTKDLKANYLSQALGAGTVWVNTYIVFGAGSPFGSKYKMSGSGRELGEY 481
Db 421 NNTYGLAAGLFTKDLRAITVSSALQAGVWVWVNCYTMLSACQCPFGGFRKMSNGRELGRH 480
QY 482 GLQAYTEVKTVTVKVPQKN 500
Db 481 GLYEYTELKTVMKISOKN 499
```

Search completed. June 24, 2003, 10 14 28
Job time : 12 0705 secs

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

697

698

699

700

701

702

703

704

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756

757

758

759

760

761

762

763

764

765

766

767

768

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807

808

809

810

811

812

813

814

815

816

817

818

819

820

821

822

823

824

825

826

827

828

829

830

831

832

833

834

835

836

837

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868

869

870

871

872

873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919

920

921

922

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

967

968

969

970

971

972

973

974

975

976

977

978

979

980

981

982

983

984

985

986

987

988

989

990

991

992

993

994

995

996

997

998

999

1000


```

QY 179 KLGPALATGNVVMKVAHATPTITALYVANI:KEAGFPFGPVVNI:VPGFGPTAGAA:1ASHED 238
Db 121 KLGPALATGNVVMKVAEATPTITALYVANI:KEAGFPFGPVVNI:VPGFGPTAGAA:1ASHED 180
QY 239 VDKVAFVTSGETIGKRVIOVAGSSNLKRVITLGLGKSPNI:IMSDADMWAVEQAHPALFFN 298
Db 181 VDKVAFVTSGETIGKRVIOVAGSSNLKRVITLGLGKSPNI:IMSDADMWAVEQAHPALFFN 240
QY 299 QGCCAGSRTFVOEDIDYDEFVRSVARAKSVVGNPFDSKTEQGPQVDETFQFKILGYI 358
Db 241 QGCCAGSRTFVOEDIDYDEFVRSVARAKSVVGNPFDSKTEQGPQVDETFQFKILGYI 300
QY 359 NTCKQEGAKLTCGGGTAADRGYFIQPTVFGDVODGNT:AKEEIFGPMVQIILKFKTIERVV 418
Db 301 KSGQEGAKLTCGGGTAADRGYFIQPTVFGDVODGNT:AKEEIFGPMVQIILKFKTIERVV 360
QY 419 GRANNSTYGLAAAVFTKLDLKNYLSQALQAGTVWVNGYDVPFAGSPGSGYKMSGSGREL 478
Db 361 GRANNSTYGLAAAVFTKLDLKNYLSQALQAGTVWVNGYDVPFAGSPGSGYKMSGSGREL 420
QY 479 GEYGLQAYTEVKTVTVKVPQKN 500
Db 421 GEYGLQAYTEVKTVTVKVPQKN 442

RESULT 2
Q8CQ2S1
ID Q8CQ2S1 PRELIMINARY: PRT: 516 AA.
AC Q8CQ2S1
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Aldehyde dehydrogenase 2.
GN ALDH2.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Taniguchi R.L., Vasilou V.:
RT "Expression and characterization of mitochondrial aldehyde
RL dehydrogenase (ALDH2) in zebrafish (Danio rerio).";
RL submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF260121; AAC19352.1;
SQ SEQUENCE 516 AA; 56562 MW; EA7D873FD6905821 CRC64;

Query Match 79.9%; Score 2097; DB 13; Length 516;
Best Local Similarity 77.8%; Pred No 1 7e-161;
Matches 389; Conservative 53; Mismatches 58; Indels 0; Gaps 0;

QY 1 MSAATQAVAPNPOPEVCNOIFINNEWHDAVSRKTEPTVNPSTGEVICOVARGKEDV 60
Db 16 ISSQHSHTIPAPNVQPDVHYNKIFINNEWHDAVSRKTEPTINPATAEVICVAGGRADV 75
QY 61 DKAREGRGAFQIGSPWRMDASHGRLNRLADLIERDRTYLALETDLONGKPYVLSYL 120
Db 76 DRAVKAARDAPFLGSSWRPMDASQGLLLNPLAQCIEPDAAYLAFTETLNGKPYTISPT 135
QY 121 VLDLWVLKCLRYACWADKYHCKTIPIDGDFEFSTRHPGVCCQIIPWNPFLLMQAKL 180
Db 136 VDLPMVVKCLRYACWADKWEKTIPIDGNVFCYTRIEPIGCVGQIIPWNPFLLMQAKL 195
QY 181 GPALATGNVVMKVAHATPTITALYVANI:KEAGFPFGPVVNI:VPGFGPTAGAA:1ASHED 240
Db 196 GPALATGNVVMKVAHATPTITALYVANI:KEAGFPFGPVVNI:VPGFGPTAGAA:1ASHED 255
QY 241 KVAFTGSTEIGKRVIOVAGSSNLKRVITLGLGKSPNI:IMSDADMWAVEQAHPALFFNQ 300
Db 256 KVAFTGSTDVGHILQOQASAKNVSLGLGKSPNI:ILSDANMEFAVEQAHPALFFNQ 315

```

```

QY 301 QGCCAGSRTFVOEDIDYDEFVRSVARAKSVVGNPFDSKTEQGPQVDETFQFKILGYINI 460
Db 316 QGCCAGSRTFVOEDIDYDEFVRSVARAKSVVGNPFDSKTEQGPQVDETFQFKILGYISS 475
QY 361 GKQEGAKLTCGGGTAADRGYFIQPTVFGDVODGNT:AKEEIFGPMVQIILKFKTIERVV 420
Db 376 GKQEGAKLTCGGGTAADRGYFIQPTVFGDVODGNT:AKEEIFGPMVQIILKFKTIERVV 435
QY 421 ANNSTYGLAAAVFTKLDLKNYLSQALQAGTVWVNGYDVPFAGSPGSGYKMSGSGREL 480
Db 436 ANNSTYGLAAAVFTKLDLKNYLSQALQAGTVWVNGYDVPFAGSPGSGYKMSGSGREL 495
QY 481 YGLQAYTEVKTVTVKVPQKN 500
Db 496 YGLQAYTEVKTVTVKVPQKN 515

RESULT 3
Q8CZS1
ID Q8CZS1 PRELIMINARY: PRT: 519 AA.
AC Q8CZS1
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 21, Last annotation update)
DE 2700007F14Rik protein (Aldehyde dehydrogenase) family, member
DE H1).
GN ALDH1B1 OR 2700007F14Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurodonta; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamahata T.,
RA Saito I., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito K.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischnann W., Gaasterland T., Gissi C., King R., Korchova H.,
RA Kuehl P., Lewis S., Matsuo Y., Nakai T., Pesole G., Quackenbush J.,
RA Sakai L.M., Staubli F., Suzuki K., Tomita M., Wagner L., Washio T.,
RA Blake J., Boffelli D., Bojunga N., Carlini P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Holman M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombauris P.,
RA Nordone P., Ring H., Ringwald M., Rodriguez L., Sakamoto N.,
RA Suzuki H., Sato K., Schenbach C., Seya T., Shibata Y., Storey K. F.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RN Nature 409:685-690(2001).
RP SEQUENCE FROM N.A.
RA Tissue-Colon;
RA Strausberg R.;
RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC 1- SIMILARITY: RELATES TO THE ALDEHYDE DEHYDROGENASES FAMILY.
DR EMBL: AK012214; BA028101.1;
DR EMBL: BC020001; AA020001.1;
DR HSP: P05091; LCW3.
DR MCD: MGI:191785; Aldh1b1.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: Pf00171; aldehyd.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLD; 1.
DR Oxidoreductase.
SQ SEQUENCE 519 AA; 57552 MW; 6322217409BEEF3 CRC64;

Query Match 74.5%; Score 1956; DB 11; Length 519;

```

[illegible]

Query Match 74.0%; Score 1943; DB 4; Length 517;
Best Local Similarity 73.1%, Proj. No. 5, 2-149,
Matches 365; Conservative 61; Mismatches 71; Indels 2; Gaps 1;

[illegible]

QY 2 SAAKQVAFAPNQCFEFCNQIEINNEBDAVSKSFTPTVNFSTGEVTCOVAGSDREYD 51
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 Db 29 SSAA--ALFSTLTNPQYNQLPINNEQDAVSKSFTPTVNFSTGEVTCGVHAFGDAIWD 77

Q7	62	KAPFGPCATGACGACGPPMPAASHGPTINPIADITFDDPTVLAALFTTIDNCXKPVVISYL	121
		: : : : : : : : : : : :	
Db	78	PAYKAAPAPPLGSGPWPMPMAASGTPLLNLIADIVFDDPTVLALETLLNGKKPFQESYAL	137
Q9	122	DLDMVLKCLKYYAGWAIRKYHGKLIPIIDGDFFSYIRKHFPGVCGGQITPWNPELLMQAKLG	181

Db 138 DLDEVIVKRYFAGWAKHKGKTIIPMDGQHCFTHREPVGVCGQIIPWNPFLVMGCKMLA 197
 QY 182 PALATGNVVMVKVAEOTPLTALYVANLKEAGFPPGVNIVPGFPTAGAAATASHEDYDK 241
 Db 198 PALATGNVVMVKVAEOTPLTALYVANLKEAGFPPGVNIVPGFPTAGAAATASHEDYDK 257
 QY 242 VAFGTSTGTEICRVIVQVAAAGSNLKRVTLELCKSPNIDMSDADMDWAVEQAIFALFFNGQ 301
 Db 258 VAFGTSTGTEVGHLLTQKAAAGDSNLRKVTLEUGKSPSIVTADAMEHAVEQHEALFFNGQ 317
 QY 302 CCAGSGSTFVQEDIDYDEFVRSVARAKSVWGNPDSKTEQGPQVDETQFKKILCYLNTG 361
 Db 318 CCAGSGSTFVQEDIDYDEFVRSVARAKSVWGNPDSKTEQGPQVDETQFKKILCYLNTG 377
 QY 362 KQSCAKILGSGTIAANDPGYFTQPTVPSVWQDMTIAKEEIEGPPVMQILKFKTIEVVGRA 421
 Db 378 QKQKALLCGRFRFGSGRFTIKPTVPGSVQVWMPVIAKEEIEGPPVMQILKFKTIEVVGRA 437
 QY 422 NSTSYGLAAAVFTKDLKKNYLSQALQAGTVWNTYVWFGAUSPRGYSKMSGSGRELGEY 481
 Db 438 NNTSYGLAAAVFTKDLKKNYLSQALQAGTVWNTYVWFGAUSPRGYSKMSGSGRELGEY 497
 QY 482 GLQAYTEVKTVTKVPQKN 500
 Db 498 GLKAYTEVKTVTKVPQKN 516

RESULT 6

Q9VLC5
 1D Q9VLC5 PRELIMINARY, PRT; 520 AA.
 AC Q9VLC5;
 DT 01-MAY-2000 (TREMblrel, 13, Created)
 DT 01-MAY-2000 (TREMblrel, 13, Last sequence update)
 DT 01-MAR-2002 (TREMblrel, 20, Last annotation update)
 DE CG3752 protein.
 GN CG3752.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota, Metazoa, Arthropoda, Tracheata, Hexapoda, Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea, Drosophilidae, Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 FX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Galinier S.E., Holt P.A., Evans C.A., Gwynne D.B.
 RA Ananadides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.F., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Rondon P.C., Peters J.H., Blazej P.G., Chang M., Pfeiffer R.P.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Anril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Berens P.V., Berland R.P., Blandstein D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Curtis J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doop L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C., Ferraz C., Ferrier S., Fleischmann W.,
 RA Fodor C., Gabriellian A.E., Garq N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Hartin D., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwu C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris T., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.P., Pauley J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

KA Reinert K., Remington K., Saunders K.D.C., Schechter F., Shen B.,
 KA Shue H.C., Siden-Kiamos L., Simpson M., Skupski M.P., Smith L.,
 KA Spier E., Spradling A.C., Stapleton M., Strout R., Sun E.,
 PA Swirskas R., Tector C., Turner R., Venter E., Ward A.H., Ward X.,
 PA Wang Z.-Y., Wasserman D.A., Weisbach C.M., Weissbach J.,
 PA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 Ye J., Yeh K.F., Zayori J.S., Zhu M., Zhang G., Zhao Q., Zhou L.,
 KA Zhou X.H., Zhou F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith B.G.,
 KA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 DR EMBL: AF003624; AAF52769.1; ...
 DR HSP: P05091; ICW3.
 DR FlyBase: Fbm0032114; CG3752.
 DR InterPro: IPR002086; Ald-lyde_dehydr.
 DR Pfam: PF00171; aldehyd; 1.
 DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLD; 1.
 KW Oxidoreductase.
 SQ SEQUENCE 520 AA; 57019 MW; 82HF54D574EC77B CR64;
 Query Match 71.1%; Score 1866; DR 5; Length 520;
 Best Local Similarity 69.8%; Pred. No. 9, 3e-14;
 Matches 351; Conservative 61; Mismatches 84; Indels 8; Gaps 4;
 QY 2 SAAATQVAPNCOPEVFCNQIFINNEWHDAVSRKTPPTVNTSGEVICQVAGSGKELVD 61
 Db 21 AVANYSSLPQPTTIDLYTVGFINNEWHKSKCKPFTINFTAEVIAETQCAKREID 80
 QY 62 KAREGHTAFQAGSGWKKMIASHSGRLKLALEDERKTYLAALFELNGKPPVIVSYV 121
 Db 81 IAVQAAPNAFKLGSPWPMVASEPGRFLYALIMEPVAVVLAETFLNCKPKPSMSYV 140
 QY 122 ELIMVLELLEPYTACWALFYHLEHLEHLEHLEHLEHLEHLEHLEHLEHLEHLEHLEH 181
 Db 141 LELTALNLEFYAGWAGNCKLEHLEHLEHLEHLEHLEHLEHLEHLEHLEHLEHLEH 200
 QY 182 PALATGNVVMVKVAEOTPLTALYVANLKEAGFPPGVNIVPGFPTAGAAATASHEDYDK 241
 Db 201 PALATGNVVMVKVAEOTPLTALYVANLKEAGFPPGVNIVPGFPTAGAAATASHEDYDK 259
 QY 242 VAFGTSTGTEICRVIVQVAAAGSNLKRVTLELCKSPNIDMSDADMDWAVEQAIFALFFNGQ 301
 Db 260 VAFGTSTGTEVGHLLTQKAAAGDSNLRKVTLEUGKSPSIVTADAMEHAVEQHEALFFNGQ 319
 QY 302 CCAGSGSTFVQEDIDYDEFVRSVARAKSVWGNPDSKTEQGPQVDETQFKKILCYLNTG 361
 Db 320 CCAGSGSTFVQEDIDYDEFVRSVARAKSVWGNPDSKTEQGPQVDETQFKKILCYLNTG 379
 QY 362 KQSCAKILGSGTIAANDPGYFTQPTVPSVWQDMTIAKEEIEGPPVMQILKFKTIEVVGRA 421
 Db 380 KQSCAKILGSGTIAANDPGYFTQPTVPSVWQDMTIAKEEIEGPPVMQILKFKTIEVVGRA 436
 QY 422 NSTSYGLAAAVFTKDLKKNYLSQALQAGTVWNTYVWFGAUSPRGYSKMSGSGRELGEY 481
 Db 437 IFRANNSEYGLAAAVFTKDLKKNYLSQALQAGTVWNTYVWFGAUSPRGYSKMSGSGRELGEY 496
 QY 478 LGEYGLQAYTEVKTVTKVPQKN 500
 Db 497 NGEYALSNTEVKSIVIVKVALN 519
 AC Q9DEX5 PWELIMINARY; PRT; 518 AA.
 ID Q9DEX5
 DT 01-MAR-2001 (TREMblrel, 16, Created)
 DT 01-MAR-2001 (TREMblrel, 16, Last sequence update)
 DT 01-MAR-2002 (TREMblrel, 21, Last annotation update)
 DE RALD12.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota, Metazoa, Chordata, Craniata; Vertebrata; Batrachostomi;


```

DE Aldehyde dehydrogenase (EC 1.2.1.3).
GN ALDH1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
KN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95191703; PubMed=10091603;
RA Ang H.L.; Duester G.;
RT *Stimulation of premature retinoic acid synthesis in Xenopus embryos
RT following premature expression of aldehyde dehydrogenase ALDH1.*;
RL Eur. J. Biochem. 260:227-234(1999)
CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
DR EMBL: AF061819; ALDH1.
DR HSSP: P51977; 1bX5.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldedh; 1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase.
SQ SEQUENCE 502 AA; 55205 MW; 54646 THRES:AA716; Ck64.

Query Match 68.3%; Score 1793; DB 13; Length 502;
Best Local Similarity 67.1%; Pred. No. 7.4e-137;
Matches 330; Conservative 67; Mismatches 95; Indels 0; Gaps 0;

QY 9 VPAPNOQPEVFCNQIFINNEWHIDVSRKTFPTVNESTGEVTCQVAGKEDVDKAREGRP 68
DB 10 LPPLLSNLEIKYTKIFINNEWHINSISGKFFPYNPATGKCEVEEGKDFVKNVAKAAR 69
QY 69 GAFGLGSPWRPMASHGRLNKLAFLEPDTYLAALFLEPDTYLAALFLEPDTYLAALF 126
DB 70 FAFGLGSPWRPMASHGRLNKLAFLEPDTYLAALFLEPDTYLAALFLEPDTYLAALF 129
QY 129 CLRYAGWADKYHCKTIPIDGDFSYTRHEPVGVGGQIIPWNPFLLMQAWKLGPAATGN 188
DB 130 SLRYCAGWADKYHCKTIPIDGDFSYTRHEPVGVGGQIIPWNPFLLMQAWKLGPAATGN 189
QY 189 VVYKVAEOTPLTALYVAVNLKEAGFPFGVNNVPGFGPTAGAAATASHEDVDKVAFTGST 248
DB 190 TVYKVAEOTPLTALYVAVNLKEAGFPFGVNNVPGFGPTAGAAATASHEDVDKVAFTGST 249
QY 249 EIGPFTQVAAAGSSNLKPVTLEGGKSNITMSDADMWAVEGAHFALEFNNQCCAGSR 308
DB 250 EVCKLKEAAGKSNLKPVTLEGGKSNITMSDADMWAVEGAHFALEFNNQCCAGSR 309
QY 309 TFVQGLIYUEFVYKSVARAKSVVGNPFDGSKTEGSPQVDETFQFKLLGYINTGQKQAGKL 368
DB 310 IFVEEPIYUEFVYKSVARAKSVVGNPFDGSKTEGSPQVDETFQFKLLGYINTGQKQAGKL 369
QY 369 LQGGTAAAGGYFLOPTVFGVQGMFTAKEEIFGVVMOILKFTIEEVGVSPANNSTYGL 428
DB 370 QCGGSAWGEKGYFSTVFSVDKQDMRLAKEELFGPVQVQILKFTIEEVGVSPANNSTYGL 429
QY 429 AAATVFKDLKANYLSOALQATVWVNYDVFGAGSPRRYKMSGSGHEIGYGLQVAYE 488
DB 430 AAGVFTKDMKATILMSALQATVWVNYDVFGAGSPRRYKMSGSGHEIGYGLQVAYE 489
QY 489 VKTIVTKVPQKN 500
DB 490 VKTIVTKVPQKN 501

RESULT 10
Q90XSR PRELIMINARY; PPT: 518 AA.
AC Q90XSR;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Retinaldehyde dehydrogenase type 2.

GN RALDH2.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
KN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21543521; PubMed=11489558;
RA Bequeman G.; Schilling T.F.; Rauch G. J.; Geisler K.; Ingham P.W.;
RT "The zebrafish neckless mutation reveals a requirement for RALDH2 in
RT mesodermal signals that pattern the hindbrain.*;
RL Development 128:1081-1094(2001).
CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
DR EMBL: AF348437; RALDH2.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldedh; 1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKN:WN.1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; UNKN:WN.2.
KW Oxidoreductase.
SQ SEQUENCE 518 AA; 56515 MW; 2060451617EHE086; Ck64.

Query Match 68.0%; Score 1786; DB 14; Length 518;
Best Local Similarity 66.9%; Pred. No. 2.9e-146;
Matches 332; Conservative 65; Mismatches 99; Indels 0; Gaps 0;

QY 4 AATQAVPAFNEQPEVFNQIFINNEWHIDVSRKTFPTVNESTGEVTCQVAGKEDVDK 64
DB 21 ASJLHSPVDPPEIKYTKIFINNEWHINSISGKFFPYNPATGKCEVEEGKDFVKNVAKAAR 80
QY 64 RESPGAGFAGGSSNKPVTLEGGKSNITMSDADMWAVEGAHFALEFNNQCCAGSR 124
DB 81 VQAAEPASFSLSASVWPMASFPCKIIFLAIVFVSALATLPSLSKFTLEPFEVDL 140
QY 124 DMVLCRLRYAGWADKYHCKTIPIDGDFSYTRHEPVGVGGQIIPWNPFLLMQAWKLGPA 184
DB 141 OGKIFTRYAGWADKYHCKTIPIDGDFSYTRHEPVGVGGQIIPWNPFLLMQAWKLGPA 200
QY 184 LATGNVVMKVAEOTPLTALYVAVNLKEAGFPFGVNNVPGFGPTAGAAATASHEDVDKVA 244
DB 201 LSGNTVVLKPAEQFTPLTALYVAVNLKEAGFPFGVNNVPGFGPTAGAAATASHEDVDKVA 260
QY 244 FTGSTEIGVTVAAAGSSNLKPVTLEGGKSNITMSDADMWAVEGAHFALEFNNQCCAGSR 304
DB 261 FTGSTEIGVTVAAAGSSNLKPVTLEGGKSNITMSDADMWAVEGAHFALEFNNQCCAGSR 320
QY 304 CASPTFVQGLIYUEFVYKSVARAKSVVGNPFDGSKTEGSPQVDETFQFKLLGYINTGQ 364
DB 321 TAGSRIFVEEPIYUEFVYKSVARAKSVVGNPFDGSKTEGSPQVDETFQFKLLGYINTGQ 380
QY 364 EGAKLGGGTAADGGYFLOPTVFGVQGMFTAKEEIFGVVMOILKFTIEEVGVSPANN 424
DB 381 EGAKLGGGTAADGGYFLOPTVFGVQGMFTAKEEIFGVVMOILKFTIEEVGVSPANN 440
QY 424 STYGLAAVFTKDLKANYLSOALQATVWVNYDVFGAGSPRRYKMSGSGHEIGYGL 484
DB 441 STYGLAAVFTKDLKANYLSOALQATVWVNYDVFGAGSPRRYKMSGSGHEIGYGL 500
QY 484 QAVTEVKTIVTKVPQKN 499
DB 501 KEYTELKTIITMKMSCK 516

RESULT 11
Q90GQ1 PRELIMINARY; PPT: 518 AA.
AC Q90GQ1;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Retinaldehyde dehydrogenase 2.
GN RALDH2.
OS Brachydanio rerio (Zebrafish) (Zebra danio).

```


OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Grandel H., Dun K., Fauch G.-J., Phinn M., Piotrowski T., Houart C.,
 RA Sordino P., Kuechler A.M., Schulte-Merker S., Geisler R., Holder N.,
 RA Wilson S.W., Brand M.;
 RT "Retinoic acid signalling in the zebrafish embryo is necessary during
 RT pre-segmentation stages to pattern the anterior-posterior axis of the
 RT CNS and to induce a pectoral fin bud.";
 RT Development 0.0.0(2002).
 DR EMBL: AF288764; AAK83071.2; -;
 SQ SEQUENCE 518 AA; 56579 MW; F0677A30F77974EA CPC64;

Query Match 67.7%; Score 1777; DB 13; Length 518;
 Best Local Similarity 66.7%; Pred. No. 3.9e-135;
 Matches 331, Conservative 65, Mismatches 100, Indels 0, Gaps 0.

QY 4 AATQAVPAPNQPEVFCNQIFINNEHDAVSRKTFPTVNPSTGEVICQVARGKUEVDVKA 63
 DB 21 ASLHLMSPVNPPEIKYIKIFINNEHDSVSGVFIYNPATGEKICDVOEADKADVDKA 80
 QY 64 PEGPPAFQIGSPMPRTASHSGLINFLADLIETPTLYLALETLDNGKPYVIVL 123
 DB 81 VQAARSAFSLGVSVMKMWASEKGLKFLADLVVERDSAYLAILELSDSGKPLPGCFVDL 140
 QY 124 DMVLKURYAGWADKYHGKTIPIGDFPSYTPHPFVGVGGQIIPNFFLLMOAKLQPA 183
 DB 141 QGIILKTRYAGWADKHGKTIPIGDFPSYTPHPFVGVGGQIIPNFFLVMTANKLQPA 200
 QY 184 LATGNVVMKVAEOTPLTALYVANLKEAGFPFGVNVNIPGPGTAAIAASHEDVDKA 243
 DB 201 LSCGNTVVLKPAEOTPLTCLYLGALKEAGFPFGVNVNIPGPGTAAIASSHMGIDKA 260
 QY 244 FTGSTEGRVQVAAAGSNLKRVTLELGGKSPNIMSDADMVAQVQAHFALFFNQGCC 303
 DB 261 FTGTEGKLVQEAAGSNLKRVTLELGGKSPNIFADADFLALBQAHQGVFFNQGCC 320
 QY 304 CAGSRTFVQEDIDYDFVVRVSARAKSRVGNPDSKTEQGPQVDETFQPKILGYNTGK 363
 DB 321 TAGSPIFVEEPIYDFVVRVSARAKSRVGNPDSKTEQGPQVDETFQPKILETQSGIT 380
 QY 364 EGAKLGGGTAADRGVFIQPTVFGVQDGMTAKFPIGPMQIILKPTTFFVVPANN 423
 DB 381 EGAKLEGCGKAPATKGEFVEPTVFSNVKDHMRKAKEEIGPVQOIMKFTIEEVIERRANN 440
 QY 424 STYGLAAAVFTKDLKANYLSQALQAGTVWVNVYDVGAGSPGYSKMSGSKELGEYGL 483
 DB 441 TEYGLAAAVFTDRIKAMTISAAVQAGTVWVNVYDVGAGSPGYSKMSGSKELGEYGL 500

RESULT 12
 Q90Y03
 ID Q90Y03 PRELIMINARY; PRT; 518 AA.
 AC Q90Y03
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DE Aldehyde dehydrogenase 1A2.
 DE Hypothetical 55.1 kDa protein.
 GN F54D8.3
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tanguay R.L., Pappa A., Vasiljov V.;

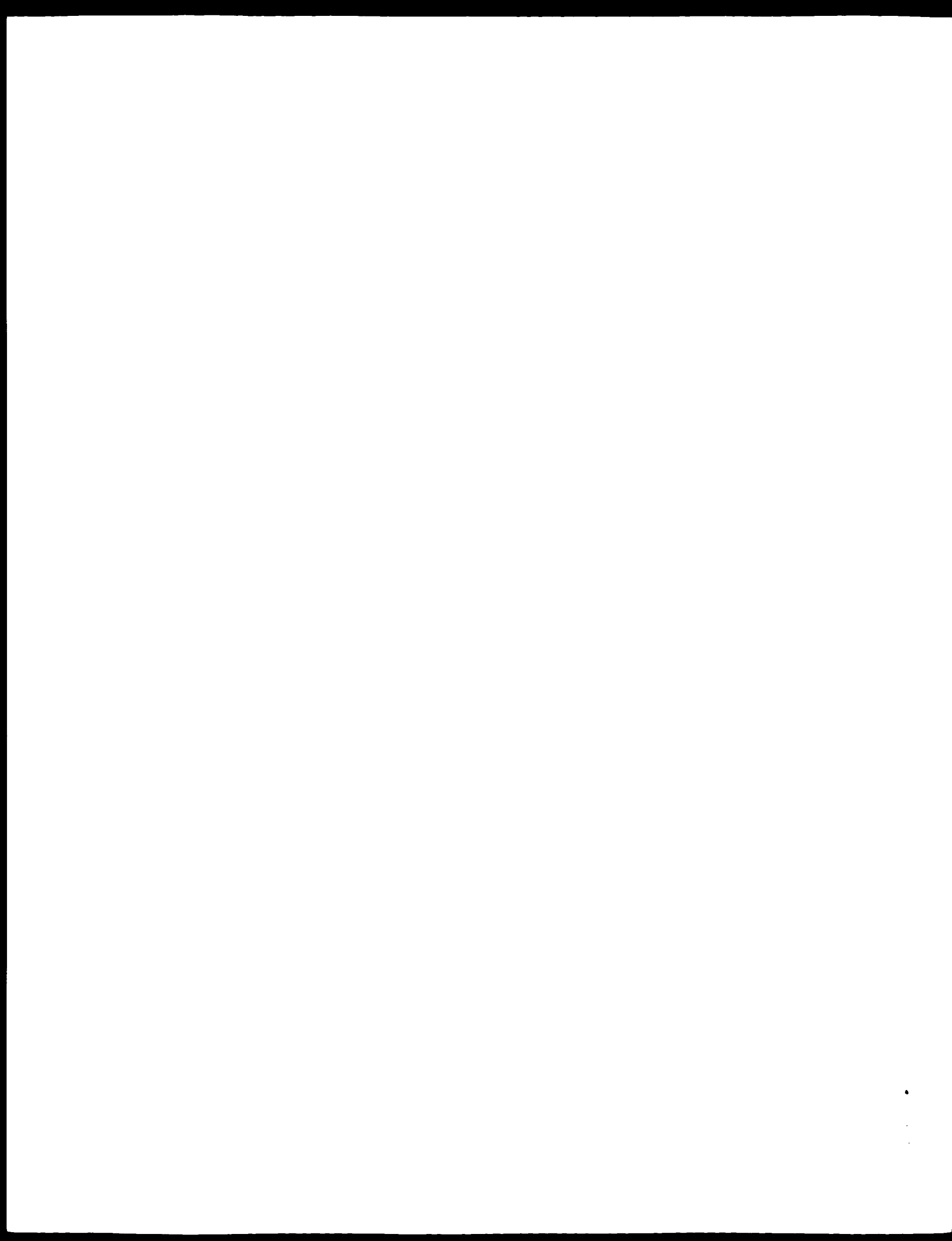
"Characterization of the zebrafish aldehyde dehydrogenase 1A2.";
 FL Submitted (01-12-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 DR EMBL: AF315631; AA:26232.1; -;
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldedh; 1.
 DR PROSITE: PS00070; ALDEHYDE_DEHYDR_GLU; UNKNOWN_1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN_2.
 KW Oxidoreductase.
 SQ SEQUENCE 518 AA; 56579 MW; A4F08D06A8P9A2R CPC64;

Query Match 67.5%; Score 1772; DB 13; Length 518;
 Best Local Similarity 66.5%; Pred. No. 3.9e-135;
 Matches 330, Conservative 65, Mismatches 101, Indels 0, Gaps 0.

QY 4 AATQAVPAPNQPEVFCNQIFINNEHDAVSRKTFPTVNPSTGEVICQVARGKUEVDVKA 63
 DB 21 ASLHLMSPVNPPEIKYIKIFINNEHDSVSGVFIYNPATGEKICDVOEADKADVDKA 80
 QY 64 PEGPPAFQIGSPMPRTASHSGLINFLADLIETPTLYLALETLDNGKPYVIVL 123
 DB 81 VQAARSAFSLGVSVMKMWASEKGLKFLADLVVERDSAYLAILELSDSGKPLPGCFVDL 140
 QY 124 DMVLKURYAGWADKYHGKTIPIGDFPSYTPHPFVGVGGQIIPNFFLLMOAKLQPA 183
 DB 141 QGIILKTRYAGWADKHGKTIPIGDFPSYTPHPFVGVGGQIIPNFFLVMTANKLQPA 200
 QY 184 LATGNVVMKVAEOTPLTALYVANLKEAGFPFGVNVNIPGPGTAAIAASHEDVDVKA 243
 DB 201 LSCGNTVVLKPAEOTPLTCLYLGALKEAGFPFGVNVNIPGPGTAAIASSHMGIDKA 260
 QY 244 FTGSTEGRVQVAAAGSNLKRVTLELGGKSPNIMSDADMVAQVQAHFALFFNQGCC 303
 DB 261 FTGTEGKLVQEAAGSNLKRVTLELGGKSPNIFADADFLALBQAHQGVFFNQGCC 320
 QY 304 CAGSRTFVQEDIDYDFVVRVSARAKSRVGNPDSKTEQGPQVDETFQPKILGYNTGK 363
 DB 321 TAGSPIFVEEPIYDFVVRVSARAKSRVGNPDSKTEQGPQVDETFQPKILETQSGIT 380
 QY 364 EGAKLGGGTAADRGVFIQPTVFGVQDGMTAKFPIGPMQIILKPTTFFVVPANN 423
 DB 381 EGAKLEGCGKAPATKGEFVEPTVFSNVKDHMRKAKEEIGPVQOIMKFTIEEVIERRANN 440
 QY 424 STYGLAAAVFTKDLKANYLSQALQAGTVWVNVYDVGAGSPGYSKMSGSKELGEYGL 483
 DB 441 TEYGLAAAVFTDRIKAMTISAAVQAGTVWVNVYDVGAGSPGYSKMSGSKELGEYGL 500

RESULT 13
 Q20780
 ID Q20780 PRELIMINARY; PRT; 510 AA.
 AC Q20780
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DE Hypothetical 55.1 kDa protein.
 GN F54D8.3
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA None;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).

Search completed: June 24, 2003, 10:17:09
Job time : 39.2679 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:42:25 ; Search time 37.544 seconds
(without alignments)
1696 348 Million cell updates/sec

Title: US-09-830-751-6

Perfect score: 2446

Sequence: 1 MSVPVQPMVDFQFVTRG.....ADKKHGLHGLATGVVVYIQS 474

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 13320620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A_Geneseq_101002.*

```

1: /SID22/qcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/qcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/qcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/qcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/qcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/qcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/qcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/qcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/qcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/qcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/qcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/qcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/qcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/qcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/qcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/qcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/qcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/qcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/qcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/qcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
21: /SID22/qcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
22: /SID22/qcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
23: /SID22/qcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2446	100.0	479	22 AAB74925	E. coli aldehyde d
2	2432	99.4	479	22 AAU29333	Novri nar re-olite
3	1407	57.5	476	21 AAY81486	Pseudomonas putida
4	932.5	38.1	493	23 ABP39543	Staphylococcus epi
5	931.5	38.1	479	22 AAC82931	S. epidermidis ope
6	879.5	36.0	483	22 AAU33562	Pseudomonas aerugi
7	873.5	35.7	482	22 AAU34671	E. coli cellular p
8	858.5	35.1	482	22 AAU38454	Salmonella typhi c
9	827	33.8	495	21 AAG51340	Arabidopsis thalia
10	827	33.8	528	21 AAG51349	Arabidopsis thalia

11	827	33.8	548	21 AAG51348	Arabidopsis thalia
12	826	33.8	495	21 AAG53558	Arabidopsis thalia
13	825	33.7	493	21 AAG53559	Arabidopsis thalia
14	811.5	32.8	488	23 ARR47807	Tristitia monocytos
15	801.5	32.8	474	22 AAR65329	P. ether 12-oxo la
16	797	32.6	509	22 ABB60875	Drosophila melano
17	796.5	32.5	442	21 AAG33550	Arabidopsis thalia
18	771.5	31.5	496	22 AAU34148	Staphylococcus aur
19	771.5	31.5	496	22 AAU37064	Staphylococcus aur
20	768.5	31.4	481	22 AAU48246	Salmonella typhi c
21	768	31.4	495	22 AAB74923	Yeast aldehyde deh
22	767.5	31.4	495	22 AAB74923	S. epidermidis ope
23	767.5	31.4	535	23 AAG33538	Staphylococcus epi
24	766.5	31.3	535	23 AAU11704	Aldehyde dehydroge
25	762.5	31.2	535	23 AAU11704	Aldehyde dehydroge
26	762.5	31.2	535	23 AAU11704	Aldehyde dehydroge
27	762	31.2	496	22 AAU34957	Enterococcus faeca
28	760.5	31.1	535	23 AAU11710	Aldehyde dehydroge
29	759.5	31.1	535	23 AAU11703	Aldehyde dehydroge
30	759.5	31.1	535	23 AAU11706	Aldehyde dehydroge
31	759.5	31.1	535	24 AAU73594	Aldehyde dehydroge
32	756.5	30.9	475	22 AAU36782	Staphylococcus aur
33	756.5	30.9	535	23 AAU11705	Aldehyde dehydroge
34	756.5	30.9	535	23 AAU11707	Aldehyde dehydroge
35	748.5	30.6	501	21 AAG36239	Arabidopsis thalia
36	748.5	30.6	501	21 AAG36239	Arabidopsis thalia
37	737.5	30.2	488	23 AAB39401	Staphylococcus epi
38	733.5	30.0	459	22 AAB22566	Protophila melano
39	729	29.8	490	22 AAG36509	Pseudomonas aerugi
40	729	29.8	490	22 AAG36509	C. glutamicum prote
41	729	29.8	490	22 AAB75349	Enterobacterium q
42	719.5	29.4	512	23 AAB51942	Prostate cancer as
43	719.5	29.4	512	23 AAB51942	Human aldehyde deh
44	719.5	29.4	529	22 AAG06577	Novri human diagno
45	719	29.4	497	7 AAB60456	Sequence of aldehy

ALIGNMENTS

RESULT 1

AAB74925

ID AAB74925 standard; Protein: 479 AA.

AC AAB74925;

ET 26-JUN-2001 (first entry)

XX E. coli aldehyde dehydrogenase aldA protein sequence SEQ ID NO:6.

DE Aldehyde dehydrogenase, glycerol dehydrogenase, 3-HP, glycerol,

XX Aldehyde dehydrogenase, glycerol dehydrogenase, 3-HP, glycerol,

XX Fungal host, 3-hydroxypropionic acid; genetic engineering, glucose;

XX Bacterial host, absorbable prosthetic device, surgical suture;

XX beta-lactam; acrylic acid; trifluoromethylated alcohol, diol,

XX polyhydroxyalkonate; copolymer; lactic acid.

XX Escherichia coli.

XX WO200116346-A1.

XX 08-MAR-2001.

PD 30-AUG-2000; 2000WO-US23878.

XX 30 AUG 1999; 93US 051440.

XX (WISC) WISCONSIN ALUMNI RES FOUND.

XX Suthers PF, Cancer DC,

XX WPI, 2001-015988/33.

XX N FSD, AAB74925.

PT 3-hydroxypropionic acid preparation, for use e.g. as monomer, by
 PT fermenting recombinant microorganisms expressing genes for suitable
 XX enzymes in the presence of glycerol or glucose -
 PS Claim 5; Page 39-41; 63pp; English.
 XX
 CC The present invention describes a method for the production of
 CC 3-hydroxypropionic acid (3-HP). The method comprises fermenting a
 CC recombinant microorganism in the presence of a source of glycerol (I)
 CC or glucose, where the microorganism: (i) expresses genes for non-native
 CC enzymes which catalyse the production of (3-HP) from (I), (ii) carries
 CC genetic constructions for the expression of a glycerol dehydratase
 CC (GDH) and aldehyde dehydrogenase (ADH) capable of catalysing the
 CC production of (3-HP) from (I); or (iii) carries a genetic construct
 CC which expresses the dhA8 gene from *Klebsiella pneumoniae* and a gene for
 CC an ADH capable of catalysing the production of (3-HP) from (I). 3-HP is
 CC a monomer, and is useful e.g. in the production of absorbable prosthetic
 CC devices and surgical sutures or for incorporation into beta-lactams,
 CC production of acrylic acid or formation of trifluoromethylated alcohols
 CC or diols, polyhydroxyalkonates and copolymers with lactic acid.
 CC Incorporation of genes encoding two enzymes makes the host organisms
 CC able to produce (3-HP) from (I). The biotechnological method of
 CC preparing (3-HP) is potentially cheaper than chemical synthesis. The
 CC present sequence represents the *E. coli* aldehyde dehydrogenase alda,
 CC which is used in the exemplification of the present invention.

XX Sequence 479 AA;

Query Match 100.0%; Score 2446; DB 22; Length 479;
 Best Local Similarity 100.0%; Pred. No. 3.6e-217;
 Matches 479; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSVPVQHPMYIDGQFVTRGDAWIDVNNIATEAVISKIPDGQAELARKAIDAAERAKQPEW 60
 DB 1 MSVPVQHPMYIDGQFVTRGDAWIDVNNIATEAVISKIPDGQAELARKAIDAAERAKQPEW 60
 QY 61 EALPATERASWLKISAGIPERASEISALIVEEGGKIQQGLAEVEVAFADYIDYMAEWAR 120
 DB 61 EALPATERASWLKISAGIPERASEISALIVEEGGKIQQGLAEVEVAFADYIDYMAEWAR 120
 QY 121 RYEGEIIQSDRPGENTILLFKRALGVTTGILPNWPFELIARKMAPALLTGNITVIKPSEF 180
 DB 121 RYEGEIIQSDRPGENTILLFKRALGVTTGILPNWPFELIARKMAPALLTGNITVIKPSEF 180
 QY 181 TTNNATAFAKIVDEIGLPRGVNVLVGRGETVQGLAGNPKVAMVSMTGSSVAGEKIMAT 240
 DB 181 TTNNATAFAKIVDEIGLPRGVNVLVGRGETVQGLAGNPKVAMVSMTGSSVAGEKIMAT 240
 QY 241 AAKNITKVCLELGGKAPAVIMDDADLELAVKALVDSRVINSQVCNCAERVVVOKGIYDQ 300
 DB 241 AAKNITKVCLELGGKAPAVIMDDADLELAVKALVDSRVINSQVCNCAERVVVOKGIYDQ 300
 QY 301 FVNRLGEAMQAVQFGNPAERNITAMGPLINAAALERVEQKVARAVEEGARVAFGKAVFG 360
 DB 301 FVNRLGEAMQAVQFGNPAERNITAMGPLINAAALERVEQKVARAVEEGARVAFGKAVFG 360
 QY 361 KGYYPPTLLLDVVRQEMSIMHEETFGPVLPPVVAFTLEDAISMANDSYGLTSSITYQNL 420
 DB 361 KGYYPPTLLLDVVRQEMSIMHEETFGPVLPPVVAFTLEDAISMANDSYGLTSSITYQNL 420
 QY 421 NVAMKALKGLKFGETVINRENFEAMQGFHAGWKKSGIGSADGKHGLHGLYLOTQVYYLQS 479
 DB 421 NVAMKALKGLKFGETVINRENFEAMQGFHAGWKKSGIGSADGKHGLHGLYLOTQVYYLQS 479

RESULT 2

AAU29333
 ID AAU29333 standard; Protein; 479 AA.

XX
 AC AAU29333;

XX
 DT 18-DEC-2001 (first entry)

XX

DE Novel mar regulated protein (NIMR) #5.
 XX mar regulated polypeptide; NIMR: microbial infection; antibacterial.
 XX Escherichia coli.
 XX W-200170776-A2.
 XX 27-SEP-2001.
 XX 38-MAR-2001; 2001W-0507478.
 XX 10-MAR-2001; 2001S-148362P.
 XX (TUFT) TUFTS COLLEGE.
 XX Levy SB, Barbosa TM, Alekshun MN;
 XX WPI: 2001-602769/68.
 XX N-PSDB: AAS46235.
 XX Identifying compounds that modulate a newly identified mar regulated
 PT polypeptide activity, useful as antimicrobial compounds, involves
 PT contacting the polypeptide with a test compound -
 XX Disclosure; Page 160-161; 526pp; English.

XX The invention relates to a method of identifying compounds that modulate
 CC a newly identified mar regulated (NIMR) polypeptide activity. The method
 CC comprises contacting an NIMR polypeptide with a test compound under
 CC interaction conditions, determining the ability of the compound to
 CC modulate the activity or expression of the polypeptide, and selecting the
 CC modulators. NIMR nucleic acids and polypeptides are used in the treatment
 CC of microbial infections, and in screening for modulators of NIMR
 CC expression and activity. These modulators can be used to reduce the
 CC infectivity of a microbe on a surface, and the virulence of a microbe in
 CC a subject suffering from an infection. AAU29333-AAU29479 represent
 CC Escherichia coli NIMR amino acid sequences of the invention.

SQ Sequence 479 AA;

Query Match 99.4%; Score 2442; DB 22; Length 479;
 Best Local Similarity 99.6%; Pred. No. 7.2e-216;
 Matches 477; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSVPVQHPMYIDGQFVTRGDAWIDVNNIATEAVISKIPDGQAELARKAIDAAERAKQPEW 60
 DB 1 MSVPVQHPMYIDGQFVTRGDAWIDVNNIATEAVISKIPDGQAELARKAIDAAERAKQPEW 60
 QY 61 EALPATERASWLKISAGIPERASEISALIVEEGGKIQQGLAEVEVAFADYIDYMAEWAR 120
 DB 61 EALPATERASWLKISAGIPERASEISALIVEEGGKIQQGLAEVEVAFADYIDYMAEWAR 120
 QY 121 RYEGEIIQSDRPGENTILLFKRALGVTTGILPNWPFELIARKMAPALLTGNITVIKPSEF 180
 DB 121 RYEGEIIQSDRPGENTILLFKRALGVTTGILPNWPFELIARKMAPALLTGNITVIKPSEF 180
 QY 181 TTNNATAFAKIVDEIGLPRGVNVLVGRGETVQGLAGNPKVAMVSMTGSSVAGEKIMAT 240
 DB 181 TTNNATAFAKIVDEIGLPRGVNVLVGRGETVQGLAGNPKVAMVSMTGSSVAGEKIMAT 240
 QY 241 AAKNITKVCLELGGKAPAVIMDDADLELAVKALVDSRVINSQVCNCAERVVVOKGIYDQ 400
 DB 241 AAKNITKVCLELGGKAPAVIMDDADLELAVKALVDSRVINSQVCNCAERVVVOKGIYDQ 400
 QY 301 FVNRLGEAMQAVQFGNPAERNITAMGPLINAAALERVEQKVARAVEEGARVAFGKAVFG 460
 DB 301 FVNRLGEAMQAVQFGNPAERNITAMGPLINAAALERVEQKVARAVEEGARVAFGKAVFG 460
 QY 361 KGYYPPTLLLDVVRQEMSIMHEETFGPVLPPVVAFTLEDAISMANDSYGLTSSITYQNL 420
 DB 361 KGYYPPTLLLDVVRQEMSIMHEETFGPVLPPVVAFTLEDAISMANDSYGLTSSITYQNL 420

Query Match 35.78; Score 873.5; DB 22; Length 482;
 Best Local Similarity 39.58; Pred. No. 8e-72;
 Matches 182; Conservative 91; Mismatches 187; Indels 1; Gaps 1.

QY 11 IDGGFTWIRGDAWIDVYNPATIAVLSIKIPGQAEDAKKALDAAEKAKPEWALPALEAS 70
 Db 15 INGEWLDANNGEADVTNPANGDKIGSVPMKGADETPAALDANRALPAWRALTAKERAT 74
 QY 71 WLKISAGIRERASEISALIVERGKIQQLAEVEAFTADYIDYMAFWAPRYEGFTQSD 130
 Db 75 ILRWFLNLMHEHDDILAKRLTLEQGRPLAEAKGEISYAAASFIEWFAEECKRIYGTITGH 134
 QY 131 RPSGNILLERKALAVTTGILPWPFPFLIARKMAPALLTGNITIVIKPSEFTTNNALATAK 190
 Db 135 QADRLIVIKQPIGVTAATTPWNPFAAMITTRKAGPALAAGCTMWLKPASQTPDSALALAE 194
 QY 191 IVDFIGLFRGVNFIIVLGRGETVQGLAGNPKVAMVSMGTGSVSGEKIMATAAKNITKVCL 250
 Db 195 LAIRAGVFAGVFNVTGSAGAVGNELTNPVLRKLSFTGSTELGRJOLMECCAKDKIKVSL 254
 QY 251 ELGGKAPAVMDADLEHAWKAIIVDSRVINSQVCMCAERYVQKGYIDQFVNRIGEAHQ 310
 Db 255 ELGNAPEFIVFDGADLEKAVESALASKFERNAGQTCVANKLYVQGVYDURFAEKLOQAVS 314
 QY 311 AVQFGNEAKNDIAMGFLINAAALERVEQKVAHVEGARVAFGKAVGKGYVYPPHIL 370
 Db 315 KLHIGDGLD NGVTIGPLIDEKAVAKVEEHIADALEKGPVVCGKAHERGGNFQPTIL 373
 QY 371 LDVQKQMSIMHEETFGPVLDPVAFDTLEDAISMANSDYGLTSSYTONLNAMKAIKGL 430
 Db 374 VDPANAKVSKETEFGPLAPLFRKDEADVIAQANDTEFLAAAYFARDLSRVFRVGEAL 433
 QY 431 KFGETYINRENFAMQGFHAGWRKSGIGGADCKHGLHGYLO 471
 Db 434 EYGVINGINTGIISNEVAPGGIKASGLRGSKYIGIEDYLE 474

RESULT 8
 AAU38454
 ID AAU38454 standard; protein: 482 AA.

AAU38454;

14-FEB-2002 (first entry)

Salmonella typhi cellular proliferation protein #345.

Antisense: prokaryotic cellular proliferation protein;

antibiotic; antibacterial; drug design.

Salmonella typhi.

W0200170955-A2.

27-SEP-2001.

21 MAR 2001; 2001W0-US09180.

21-MAR-2000; 2000US-191078P.

23-MAY-2000; 2000US-206848P.

26-MAY-2000; 2000US-207727P.

23-OCT-2000; 2000US-242578P.

27-NOV-2000; 2000US-253625P.

22-DEC-2000; 2000US-257931P.

16-FEB-2001; 2001US-259108P.

(ELIT-) ELITRA PHARM INC.

Haselbeck R, Ohlson KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

Yamamoto RT, Xu RH;

WPI; 2001-611495/70.

N-PSDH; AAS56313.

New polynucleotides for the identification and development of
 antibiotics, comprise sequences of antisense nucleic acids

Example 3; Seq ID No 14047; 511pp; English.

The invention relates to antisense inhibitors of genes essential to
 prokaryotic cellular proliferation, their use in identifying the
 genes, their use in the discovery of novel antibiotics, the essential
 genes themselves and the encoded proteins. The prokaryotes used are
 Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 pneumoniae, pseudomonas aeruginosa and Enterococcus faecalis. The
 invention is also useful for the identification of potential new targets
 for antibiotic development. The antisense nucleic acids can also be used
 to identify proteins used in proliferation, to express these proteins,
 and to obtain antibodies capable of binding to the expressed proteins.
 The proteins can be used to screen compounds in rational drug discovery
 programmes. The antisense nucleic acid sequence is also useful to screen
 for homologous nucleic acids which are required for cell proliferation in
 a wide variety of organisms. The present sequence represents an
 essential prokaryotic cellular proliferation protein.

Note: The sequence data for this patent did not form part
 of the printed specification, but was obtained in electronic
 format directly from WIPO at

ftp.wipo.int/pub/published_jpat_sequences.

Sequence 482 AA;

Query Match 35.18; Score 858.5; DB 22; Length 482;
 Best Local Similarity 39.28; Pred. No. 2e-70;
 Matches 181; Conservative 91; Mismatches 189; Indels 1; Gaps 1;

QY 10 YIDGQFVTRHGDAWIDVYNPATIAVLSIKIPGQAEDAKKALDAAEKAKPEWALPALEAS 69
 Db 14 FIDGWRDARGGQVIVPSNPANKPLGNNVFKMCAETRTDIAINAKKALPAWALTAKERA 74
 QY 70 SWLRKISAGIRERASEISALIVERGKIQQLAEVEAFTADYIDYMAFWAPRYEGFTQSD 129
 Db 74 NLLRWFLNLMHEHDDILAKRLTLEQGRPLAEAKGEISYAAASFIEWFAEECKRIYGTITGH 134
 QY 130 DRPCENILLERKALAVTTGILPWPFPFLIARKMAPALLTGNITIVIKPSEFTTNNALATAK 189
 Db 134 BQDKRLIVIKQPIGVTAATTPWNPFAAMITTRKAGPALAAGCTMWLKPASQTPDSALALAE 194
 QY 190 KIVDEIGLFRGVNFIIVLGRGETVQGLAGNPKVAMVSMGTGSVSGEKIMATAAKNITKVCL 249
 Db 194 ELAQBAGILLAGVFNVTGSAGAVGNELTNPVLRKLSFTGSTELGRJOLMECCAKDKIKVSL 254
 QY 250 LEUGKAPAVMDADLEHAWKAIIVDSRVINSQVCMCAERYVQKGYIDQFVNRIGEAHQ 309
 Db 254 LEUGGNAPFIVFDGADLEKAVESALASKFERNAGQTCVANKLYVQGVYDURFAEKLOQAV 314
 QY 310 QAVQFGNEAKNDIAMGFLINAAALERVEQKVAHVEGARVAFGKAVGKGYVYPPHIL 369
 Db 314 NKLAVGDLGDA IVAIGPLIDEKAVAKVEEHIADALEKGPVVCGKAHERGGNFQPTIL 472
 QY 370 LDVQKQMSIMHEETFGPVLDPVAFDTLEDAISMANSDYGLTSSYTONLNAMKAIKGL 429
 Db 373 LAWTDNRAKVAKEITFGPLAPLFRKDEADVIAQANDTEFLAAAYFARDLSRVFRVGEAL 432
 QY 430 KFGETYINRENFAMQGFHAGWRKSGIGGADCKHGLHGYLO 471
 Db 433 LEYGVINGINTGIISNEVAPGGIKASGLRGSKYIGIEDYLE 474

RESULT 9

AAG51350

ID AAG51350 standard; protein: 495 AA.

AC AAG51350;

18-OCT-2000 (first entry)

us-09-830-751-6.rag

Tue Jun 24 10:42:22 2003

XX Arabidopsis thaliana protein fragment SEQ ID NO: 65163.
DE Protein identification: signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0131875.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126284.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134221.
XX 18-MAY-1999; 99US-0134370.
XX 19-MAY-1999; 99US-0134768.
XX 20-MAY-1999; 99US-0134941.
XX 21-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 25-MAY-1999; 99US-0136392.
XX 27-MAY-1999; 99US-0136782.
XX 28-MAY-1999; 99US-0137232.
XX 01-JUN-1999; 99US-0137528.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.
XX 18-JUN-1999; 99US-0139763.
XX 21-JUN-1999; 99US-0139817.
XX 22-JUN-1999; 99US-0139899.
XX 23-JUN-1999; 99US-0140353.
XX 23-JUN-1999; 99US-0140354.
XX 24-JUN-1999; 99US-0140695.
XX 28-JUN-1999; 99US-0140823.
XX 29-JUN-1999; 99US-0140991.
XX 30-JUN-1999; 99US-0141287.
XX 01-JUL-1999; 99US-0141842.
XX 01-JUL-1999; 99US-0142154.
XX 02-JUL-1999; 99US-0142055.
XX 06-JUL-1999; 99US-0142390.
XX 08-JUL-1999; 99US-0142803.
XX 09-JUL-1999; 99US-0142920.
XX 12-JUL-1999; 99US-0142977.
XX 13-JUL-1999; 99US-0143542.
XX 14-JUL-1999; 99US-0143624.
XX 15-JUL-1999; 99US-0144005.
XX 16-JUL-1999; 99US-0144085.
XX 16-JUL-1999; 99US-0144086.
XX 19-JUL-1999; 99US-0144325.
XX 19-JUL-1999; 99US-0144331.
XX 19-JUL-1999; 99US-0144332.
XX 19-JUL-1999; 99US-0144333.
XX 19-JUL-1999; 99US-0144335.
XX 20-JUL-1999; 99US-0144352.
XX 20-JUL-1999; 99US-0144632.
XX 20-JUL-1999; 99US-0144884.
XX 21-JUL-1999; 99US-0144814.
XX 21-JUL-1999; 99US-0145086.
XX 21-JUL-1999; 99US-0145088.
XX 22-JUL-1999; 99US-0145085.
XX 22-JUL-1999; 99US-0145087.
XX 22-JUL-1999; 99US-0145089.
XX 22-JUL-1999; 99US-0145192.
XX 23-JUL-1999; 99US-0145145.
XX 23-JUL-1999; 99US-0145218.
XX 23-JUL-1999; 99US-0145224.
XX 26-JUL-1999; 99US-0145276.
XX 27-JUL-1999; 99US-0145913.
XX 27-JUL-1999; 99US-0145918.
XX 27-JUL-1999; 99US-0145919.
XX 28-JUL-1999; 99US-0145951.
XX 02-AUG-1999; 99US-0146386.
XX 02-AUG-1999; 99US-0146388.
XX 02-AUG-1999; 99US-0146389.
XX 03-AUG-1999; 99US-0147038.
XX 04-AUG-1999; 99US-0147204.
XX 04-AUG-1999; 99US-0147302.
XX 05-AUG-1999; 99US-0147240.
XX 05-AUG-1999; 99US-0147240.
XX 06-AUG-1999; 99US-0147303.
XX 06-AUG-1999; 99US-0147416.
XX 09-AUG-1999; 99US-0147493.
XX 09-AUG-1999; 99US-0147935.
XX 09-AUG-1999; 99US-0148171.
XX 10-AUG-1999; 99US-0148319.
XX 12-AUG-1999; 99US-0148341.
XX 13-AUG-1999; 99US-0148565.
XX 13-AUG-1999; 99US-0148684.
XX 16-AUG-1999; 99US-0149368.
XX 17-AUG-1999; 99US-0149175.
XX 18-AUG-1999; 99US-0149426.
XX 18-AUG-1999; 99US-0149722.
XX 20-AUG-1999; 99US-0149723.
XX 20-AUG-1999; 99US-0149929.
XX 23-AUG-1999; 99US-0149902.
XX 23-AUG-1999; 99US-0149930.
XX 25-AUG-1999; 99US-0150566.
XX 26-AUG-1999; 99US-0150884.
XX 27-AUG-1999; 99US-0151065.
XX 27-AUG-1999; 99US-0151066.

PR 27-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137502.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 10-JUN-1999; 9905-0138847.
PR 14-JUN-1999; 9905-0139119.
PR 16-JUN-1999; 9905-0139452.
PR 16-JUN-1999; 9905-0139453.
PR 17-JUN-1999; 9905-0139492.
PR 18-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139763.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139899.
PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 01-JUL-1999; 9905-0142154.
PR 02-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142490.
PR 06-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142920.
PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143544.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144632.
PR 20-JUL-1999; 9905-0144684.
PR 21-JUL-1999; 9905-0144814.
PR 21-JUL-1999; 9905-0144814.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.
PR 22-JUL-1999; 9905-0145085.
PR 22-JUL-1999; 9905-0145087.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145132.
PR 23-JUL-1999; 9905-0145132.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.
PR 23-JUL-1999; 9905-0145224.
PR 26-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145913.
PR 27-JUL-1999; 9905-0145918.
PR 27-JUL-1999; 9905-0145919.
PR 28-JUL-1999; 9905-0145951.
PR 02-AUG-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146388.
PR 02-AUG-1999; 9905-0146389.
PR 03-AUG-1999; 9905-0147049.
PR 04-AUG-1999; 9905-0147204.
PR 04-AUG-1999; 9905-0147402.
PR 05-AUG-1999; 9905-0147432.
PR 05-AUG-1999; 9905-0147460.
PR 06-AUG-1999; 9905-0147403.
PR 06-AUG-1999; 9905-0147416.
PR 06-AUG-1999; 9905-0147493.
PR 06-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148171.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148565.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149922.
PR 23-AUG-1999; 9905-0149940.
PR 25-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151403.
PR 31-AUG-1999; 9905-0151438.
PR 01-SEP-1999; 9905-0151830.
PR 07-SEP-1999; 9905-0152363.
PR 10-SEP-1999; 9905-0153070.
PR 13-SEP-1999; 9905-0153758.
PR 15-SEP-1999; 9905-0154018.
PR 16-SEP-1999; 9905-0154039.
PR 20-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155133.
PR 23-SEP-1999; 9905-0155486.
PR 24-SEP-1999; 9905-0155659.
PR 28-SEP-1999; 9905-0156458.
PR 29-SEP-1999; 9905-0156596.
PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157753.
PR 06-OCT-1999; 9905-0157865.
PR 07-OCT-1999; 9905-0158029.
PR 08-OCT-1999; 9905-0158232.
PR 12-OCT-1999; 9905-0158469.
PR 13-OCT-1999; 9905-0159293.
PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159395.
PR 14-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 14-OCT-1999; 9905-0159684.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160768.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160814.
PR 21-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160980.
PR 22-OCT-1999; 9905-0160981.
PR 25-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161406.
PR 25-OCT-1999; 9905-0161584.
PR 25-OCT-1999; 9905-0161460.
PR 25-OCT-1999; 9905-0161361.
PR 25-OCT-1999; 9905-0161400.
PR 28-OCT-1999; 9905-0161592.
PR 28-OCT-1999; 9905-0161993.


```
PR 19-JUL-1999; 99US-0144333
PR 19-JUL-1999; 99US-0144334
PR 19-JUL-1999; 99US-0144335
PR 20-JUL-1999; 99US-0144352
PR 20-JUL-1999; 99US-0144632
PR 20-JUL-1999; 99US-0144884
PR 21-JUL-1999; 99US-0144814
PR 21-JUL-1999; 99US-0145086
PR 21-JUL-1999; 99US-0145088
PR 21-JUL-1999; 99US-0145085
PR 22-JUL-1999; 99US-0145087
PR 22-JUL-1999; 99US-0145089
PR 22-JUL-1999; 99US-0145192
PR 23-JUL-1999; 99US-0145145
PR 23-JUL-1999; 99US-0145218
PR 23-JUL-1999; 99US-0145224
PR 26-JUL-1999; 99US-0145276
PR 27-JUL-1999; 99US-0145913
PR 27-JUL-1999; 99US-0145918
PR 27-JUL-1999; 99US-0145919
PR 28-JUL-1999; 99US-0145951
PR 02-AUG-1999; 99US-0145384
PR 02-AUG-1999; 99US-0145388
PR 02-AUG-1999; 99US-0145389
PR 03-AUG-1999; 99US-0147038
PR 04-AUG-1999; 99US-0147204
PR 04-AUG-1999; 99US-0147302
PR 05-AUG-1999; 99US-0147192
PR 05-AUG-1999; 99US-0147260
PR 06-AUG-1999; 99US-0147403
PR 06-AUG-1999; 99US-0147416
PR 09-AUG-1999; 99US-0147493
PR 09-AUG-1999; 99US-0147935
PR 10-AUG-1999; 99US-0148171
PR 11-AUG-1999; 99US-0148319
PR 12-AUG-1999; 99US-0148341
PR 13-AUG-1999; 99US-0148565
PR 13-AUG-1999; 99US-0148684
PR 16-AUG-1999; 99US-0149368
PR 17-AUG-1999; 99US-0149175
PR 18-AUG-1999; 99US-0149426
PR 20-AUG-1999; 99US-0149723
PR 20-AUG-1999; 99US-0149723
PR 20-AUG-1999; 99US-0149723
PR 23-AUG-1999; 99US-0149930
PR 23-AUG-1999; 99US-0149930
PR 25-AUG-1999; 99US-0150566
PR 26-AUG-1999; 99US-0150884
PR 27-AUG-1999; 99US-0151065
PR 27-AUG-1999; 99US-0151065
PR 27-AUG-1999; 99US-0151080
PR 30-AUG-1999; 99US-0151303
PR 31-AUG-1999; 99US-0151338
PR 01-SEP-1999; 99US-0151940
PR 07-SEP-1999; 99US-0152363
PR 10-SEP-1999; 99US-0153070
PR 13-SEP-1999; 99US-0153758
PR 13-SEP-1999; 99US-0154018
PR 16-SEP-1999; 99US-0154039
PR 20-SEP-1999; 99US-0154779
PR 22-SEP-1999; 99US-0155139
PR 23-SEP-1999; 99US-0155486
PR 24-SEP-1999; 99US-0155659
PR 28-SEP-1999; 99US-0156458
PR 29-SEP-1999; 99US-0156596
PR 04-OCT-1999; 99US-0157117
PR 05-OCT-1999; 99US-0157753
PR 06-OCT-1999; 99US-0157865
PR 07-OCT-1999; 99US-0158029
PR 08-OCT-1999; 99US-0158232
PR 12-OCT-1999; 99US-0158369
PR 13-OCT-1999; 99US-0159293
PR 13-OCT-1999; 99US-0159294
```

```
PR 13-OCT-1999; 99US-0159295
PR 14-OCT-1999; 99US-0159329
PR 14-OCT-1999; 99US-0159330
PR 14-OCT-1999; 99US-0159331
PR 14-OCT-1999; 99US-0159637
PR 14-OCT-1999; 99US-0159638
PR 18-OCT-1999; 99US-0159584
PR 21-OCT-1999; 99US-0160741
PR 21-OCT-1999; 99US-0160767
PR 21-OCT-1999; 99US-0160768
PR 21-OCT-1999; 99US-0160770
PR 21-OCT-1999; 99US-0160814
PR 21-OCT-1999; 99US-0160815
PR 22-OCT-1999; 99US-0160980
PR 22-OCT-1999; 99US-0160981
PR 22-OCT-1999; 99US-0160981
PR 25-OCT-1999; 99US-0161404
PR 25-OCT-1999; 99US-0161405
PR 25-OCT-1999; 99US-0161406
PR 26-OCT-1999; 99US-0161359
PR 26-OCT-1999; 99US-0161360
PR 26-OCT-1999; 99US-0161361
PR 26-OCT-1999; 99US-0161920
PR 28-OCT-1999; 99US-0161992
PR 28-OCT-1999; 99US-0161993
PR 29 OCT 1999; 99US 0162142

Query Match 33.8%; Score 827; DB 21; Length 548;
Best Local Similarity 37.8%; Pred. No. 1.9e-67;
Matches 185; Conservative 90; Mismatches 194; Indels 22; Gaps 3;

9 MYIDQFVWTP-----GDWID-----VVPATEAVISIPGQADAP 47
54 MSMPALSVSEKLPSSGLLPFLDQKWLISYFNKTIKVNPNATGEIADVACMGKETS 113
48 KATDAEPAQDPEWALPAIEPASWIRKISAGTIPPSASIALIVFSGKIQQLAFVPAF 107
114 DATASSYEAFTSWSPPLTAGESKVLPPWYNLLIAHKEFLGQLITLFGSKPLKFAIGVAY 173
108 TADYIYMAEWAPYEGPTIGSTPRRENITLIFKALGVTTGTLFWNPFFLIAPKNAPAL 167
174 GASFLEYALLAKPVYGLITFNLSDFPLVIGFVGVCAITWTFPLAMITKVKCFAL 233
108 LGNTIVKISLFTINNAIAFAKIVBELGLIGOVENIVLGRCELVGLEAGNFKVAVSM 227
234 ASPTVWVVEFSPLFTIATAAARLACAGVPTGALNVMMENADPTFPAITTSQVPPITF 293
238 ICSVSAGKIMAJAAKNILKVLELGKKAFAVMUAWGLFVAVFVSVAVHSQAVGNC 287
294 TGSTAVGPKLMAAAAFVVERVSTFISNAISIVTFDAGGVAVVGTIAKFRSQQLQVC 353
288 APPVYVVKGTYNQFVNPICTEAMQAVQFNAPFNDIANGDTINAAALPPVAFVPAVVEE 347
354 ANPLVLDGIVKFAEAFSEAVGKIFVGIQGFPTTGGPIINAAVQKVEIFVQAVSK 412
348 GAPVAFGKAVGKGYYPPTLLLDVROEMSIMHEETPGVLPVPAEDTLBDALSMANDS 407
413 GAKIIGGKPHSLGNTFYPTVTPVSDNMIMSKEEIEFGVAPLIPKTFEDAIPLANIT 472
408 DYGLISSIYTONLNVAMKAIKGLAKGETYINKENFEAMQGHGAKPKSGIGGAGKHGLH 467
473 IAGLAAYIFTNVSQESWVFEALLEYELGVGNFELISTFVAFPGVAFVSGTGPFGSGYGM 542
468 GYLQTVVYL 477
533 EYLEIKYVCL 542

RESULT 12
AAC23358
ID AAG23558 standard; Protein; 495 AA.
XX
AC AAC23558;
```

XX 17-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 26911.
DE
XX
XX Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 990S-0121825.
XX 05-MAR-1999; 990S-0123180.
XX 09-MAR-1999; 990S-0123548.
XX 23-MAR-1999; 990S-0125788.
XX 25-MAR-1999; 990S-0126264.
XX 29-MAR-1999; 990S-0126785.
XX 01-APR-1999; 990S-0127462.
XX 06-APR-1999; 990S-0128234.
XX 08-APR-1999; 990S-0128714.
XX 16-APR-1999; 990S-0129845.
XX 19-APR-1999; 990S-0130077.
XX 21-APR-1999; 990S-0130449.
XX 23-APR-1999; 990S-0130510.
XX 23-APR-1999; 990S-0130894.
XX 28-APR-1999; 990S-0131449.
XX 30-APR-1999; 990S-0132048.
XX 30-APR-1999; 990S-0132407.
XX 04-MAY-1999; 990S-0132484.
XX 05-MAY-1999; 990S-0132485.
XX 06-MAY-1999; 990S-0132486.
XX 06-MAY-1999; 990S-0132487.
XX 07-MAY-1999; 990S-0132863.
XX 11-MAY-1999; 990S-0134256.
XX 14-MAY-1999; 990S-0134218.
XX 14-MAY-1999; 990S-0134219.
XX 14-MAY-1999; 990S-0134221.
XX 14-MAY-1999; 990S-0134370.
XX 18-MAY-1999; 990S-0134768.
XX 19-MAY-1999; 990S-0134941.
XX 20-MAY-1999; 990S-0135124.
XX 21-MAY-1999; 990S-0135453.
XX 24-MAY-1999; 990S-0135629.
XX 25-MAY-1999; 990S-0136021.
XX 27-MAY-1999; 990S-0136392.
XX 28-MAY-1999; 990S-0136782.
XX 01-JUN-1999; 990S-0137222.
XX 03-JUN-1999; 990S-0137528.
XX 04-JUN-1999; 990S-0137502.
XX 07-JUN-1999; 990S-0137724.
XX 08-JUN-1999; 990S-0138094.
XX 10-JUN-1999; 990S-0138540.
XX 10-JUN-1999; 990S-0138847.
XX 14-JUN-1999; 990S-0139119.
XX 16-JUN-1999; 990S-0139452.
XX 16-JUN-1999; 990S-0139453.
XX 17-JUN-1999; 990S-0139492.
XX 18-JUN-1999; 990S-0139454.
XX 18-JUN-1999; 990S-0139455.
XX 18-JUN-1999; 990S-0139456.
XX 18-JUN-1999; 990S-0139457.
XX 18-JUN-1999; 990S-0139458.
XX 18-JUN-1999; 990S-0139459.
XX 18-JUN-1999; 990S-0139460.
XX 18-JUN-1999; 990S-0139461.
XX 18-JUN-1999; 990S-0139462.
XX 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139764.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139899.
PR 23-JUN-1999; 990S-0140353.
PR 23-JUN-1999; 990S-0140454.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140824.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141297.
PR 01-JUL-1999; 990S-0141842.
PR 01-JUL-1999; 990S-0142154.
PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.
PR 08-JUL-1999; 990S-0142804.
PR 09-JUL-1999; 990S-0142920.
PR 12-JUL-1999; 990S-0142977.
PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.
PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144884.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 21-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145085.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145192.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145218.
PR 24-JUL-1999; 990S-0145224.
PR 26-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145913.
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 02-AUG-1999; 990S-0146389.
PR 03-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 04-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 05-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147204.
PR 06-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 09-AUG-1999; 990S-0147935.
PR 11-AUG-1999; 990S-0148371.
PR 12-AUG-1999; 990S-0148319.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0149368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149725.
PR 20-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149930.
PR 25-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.


```
PR 27-AUG-1999; 900S-0151065
PR 27-AUG-1999; 900S-0151066
PR 27-AUG-1999; 900S-0151067
PR 30-AUG-1999; 900S-0151303
PR 30-AUG-1999; 900S-0151304
PR 30-AUG-1999; 900S-0151305
PR 01-SEP-1999; 900S-0151930
PR 01-SEP-1999; 900S-0151931
PR 01-SEP-1999; 900S-0152363
PR 07-SEP-1999; 900S-0153070
PR 07-SEP-1999; 900S-0153758
PR 13-SEP-1999; 900S-0154018
PR 15-SEP-1999; 900S-0154039
PR 16-SEP-1999; 900S-0154774
PR 20-SEP-1999; 900S-0155139
PR 22-SEP-1999; 900S-0155486
PR 23-SEP-1999; 900S-0155659
PR 24-SEP-1999; 900S-0156596
PR 28-SEP-1999; 900S-0156596
PR 29-SEP-1999; 900S-0157117
PR 04-OCT-1999; 900S-0157753
PR 05-OCT-1999; 900S-0157865
PR 06-OCT-1999; 900S-0158029
PR 07-OCT-1999; 900S-0158232
PR 08-OCT-1999; 900S-0158359
PR 12-OCT-1999; 900S-0159293
PR 13-OCT-1999; 900S-0159294
PR 13-OCT-1999; 900S-0159295
PR 13-OCT-1999; 900S-0159329
PR 14-OCT-1999; 900S-0159330
PR 14-OCT-1999; 900S-0159331
PR 14-OCT-1999; 900S-0159637
PR 14-OCT-1999; 900S-0159638
PR 18-OCT-1999; 900S-0159584
PR 21-OCT-1999; 900S-0160741
PR 21-OCT-1999; 900S-0160767
PR 21-OCT-1999; 900S-0160768
PR 21-OCT-1999; 900S-0160770
PR 21-OCT-1999; 900S-0160814
PR 21-OCT-1999; 900S-0160815
PR 22-OCT-1999; 900S-0160880
PR 22-OCT-1999; 900S-0160981
PR 22-OCT-1999; 900S-0160989
PR 25-OCT-1999; 900S-0161404
PR 25-OCT-1999; 900S-0161405
PR 25-OCT-1999; 900S-0161406
PR 26-OCT-1999; 900S-0161359
PR 26-OCT-1999; 900S-0161360
PR 26-OCT-1999; 900S-0161361
PR 28-OCT-1999; 900S-0161920
PR 28-OCT-1999; 900S-0161942
PR 28-OCT-1999; 900S-0161943
PR 29-OCT-1999; 900S-0162142

Query Match
Best Local Similarity 37.68; Pred. No. 2; Length 495;
Matches 184; Conservative 91; Mismatches 193; Indels 22; Gaps 3.

QY 9 MYDQPVWTR-----GDWIDP-----VYNPATEAVISRPDGOARAP 47
DB 1 MSMDAQSVKRSGLRQGLIGSKWILGYSYINKILKVNPAFGFIALVAWMTIKETN 60
QY 48 KALDAAPAPAEWALAIERASWIKTSACIPEPASEISALIVPGRGKIQALAEVAP 107
DB 61 DATASYEFTSWSPITAGERSKVLRFWYVULLIAHKEELQITLLEQGRPLKEATGEVAY 120
QY 108 TADYIDYMAEMARVEGETTOSORPGENILLERKALGVITIGILPWNFPFLAKKMAPAL 167
DB 121 GASFLYYAEAKKRVYGDILPPNLSDRLLVLKQPGVGTGAITPWNFPLAMITRKVSPAL 180
QY 168 LTRNTVTIKSEFTTNAIAFAKIVNPLCLPPGVNPLNLRGSETVQDELAGNFKVAMVSM 227
DB 181 ASGCTVVVKPSBELPLTALAAELAQGVPPGALNVVMGNAPELIGDALLTSPQVRKITF 240
QY 228 TCSVSAFKMTAKKNTKVCLELOSKAIAIWMDEADLELAVKAIVUSRVINSQVWV 247
```

241 TGTAVGCKKLMAAAATPTKVKVSLGONAPSVIFDDADLDVAVKGTAAKAFNSQCTVC 300

288 AEPVYVKGCIYDPNPLGFAAGAVAGPNDPAFPNDIAMGFLINAAALEPVEVAVAPAVEE 347

301 ANPVLVIGIYTFWFAFAFSEAVAKLEVAGGPPGDTIGGFLINDAAVQKVEFVQDAVSK 359

348 GAPVAFGSKAVFGKGYGYYPTTILIDVPQFMSIMHEFTFGVLPVVAEDTLEDAISMNUS 407

360 GAKIIGGKPHSLGSMTFEYPTVDPVSDNNIMSKKEIFGCVAPLIPFKTEEDAIRANIT 419

408 LYGLEISSIYDNLNVAMKALKGKPEFTYINPNEFAMLGFPHAWKFSIFGALGKHGLH 467

420 IAGLAAYLPINSVQKSWKPVFEALEYGLVAVVNEELISTFVAFPGVYKQSLGKPEFSKYMD 479

468 GYLQIQVYVL 477

480 EYLEIKYVCL 489

RESULT 13

AAC23559 standard; Protein; 493 AA.

XX AC AAC23559;

XX DE 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 26912.

KW Protein identification, signal transduction pathway, metabolic pathway, hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301479

XX 25-FEB-1999; 990S-0121825.

XX 05-MAR-1999; 990S-0123419.

XX 09-MAR-1999; 990S-0123548.

XX 23-MAR-1999; 990S-0125788.

XX 25-MAR-1999; 990S-0126264.

XX 29-MAR-1999; 990S-0126785.

XX 01-APR-1999; 990S-0127462.

XX 04-APR-1999; 990S-0128234.

XX 08-APR-1999; 990S-0128714.

XX 16-APR-1999; 990S-0129845.

XX 19-APR-1999; 990S-0130077.

XX 21-APR-1999; 990S-0130449.

XX 23-APR-1999; 990S-0130510.

XX 23-APR-1999; 990S-0130891.

XX 28-APR-1999; 990S-0131449.

XX 30-APR-1999; 990S-0132048.

XX 30-APR-1999; 990S-0132407.

XX 04-MAY-1999; 990S-0132484.

XX 04-MAY-1999; 990S-0132485.

XX 06-MAY-1999; 990S-0132486.

XX 06-MAY-1999; 990S-0132487.

XX 07-MAY-1999; 990S-0132863.

XX 11-MAY-1999; 990S-0134256.

XX 14-MAY-1999; 990S-0134218.

XX 14-MAY-1999; 990S-0134219.

XX 14-MAY-1999; 990S-0134221.

XX 14-MAY-1999; 990S-0134370.

XX 18-MAY-1999; 990S-0134758.

XX 19-MAY-1999; 990S-0134941.

XX 20-MAY-1999; 990S-0135124.

XX 21-MAY-1999; 990S-0135353.

PR 24 MAY-1999; 990S-0135629.
PR 25 MAY-1999; 990S-0136021.
PR 27 MAY-1999; 990S-0136392.
PR 28 MAY-1999; 990S-0136782.
PR 01 JUN-1999; 990S-0137222.
PR 03 JUN-1999; 990S-0137528.
PR 04 JUN-1999; 990S-0137502.
PR 07 JUN-1999; 990S-0137724.
PR 08 JUN-1999; 990S-0138094.
PR 10 JUN-1999; 990S-0138540.
PR 10 JUN-1999; 990S-0138847.
PR 14 JUN-1999; 990S-0139119.
PR 16 JUN-1999; 990S-0139452.
PR 16 JUN-1999; 990S-0139453.
PR 17 JUN-1999; 990S-0139492.
PR 18 JUN-1999; 990S-0139454.
PR 18 JUN-1999; 990S-0139455.
PR 18 JUN-1999; 990S-0139456.
PR 18 JUN-1999; 990S-0139457.
PR 18 JUN-1999; 990S-0139458.
PR 18 JUN-1999; 990S-0139459.
PR 18 JUN-1999; 990S-0139460.
PR 18 JUN-1999; 990S-0139461.
PR 18 JUN-1999; 990S-0139462.
PR 18 JUN-1999; 990S-0139463.
PR 18 JUN-1999; 990S-0139750.
PR 18 JUN-1999; 990S-0139763.
PR 21 JUN-1999; 990S-0139817.
PR 22 JUN-1999; 990S-0139899.
PR 23 JUN-1999; 990S-0140353.
PR 23 JUN-1999; 990S-0140354.
PR 24 JUN-1999; 990S-0140355.
PR 28 JUN-1999; 990S-0140695.
PR 29 JUN-1999; 990S-0140823.
PR 29 JUN-1999; 990S-0140991.
PR 30 JUN-1999; 990S-0141287.
PR 01 JUL-1999; 990S-0141842.
PR 01 JUL-1999; 990S-0142154.
PR 02 JUL-1999; 990S-0142055.
PR 06 JUL-1999; 990S-0142190.
PR 08 JUL-1999; 990S-0142803.
PR 09 JUL-1999; 990S-0142920.
PR 12 JUL-1999; 990S-0142977.
PR 13 JUL-1999; 990S-0143542.
PR 14 JUL-1999; 990S-0143624.
PR 15 JUL-1999; 990S-0144005.
PR 16 JUL-1999; 990S-0144085.
PR 16 JUL-1999; 990S-0144085.
PR 19 JUL-1999; 990S-0144325.
PR 19 JUL-1999; 990S-0144331.
PR 19 JUL-1999; 990S-0144332.
PR 19 JUL-1999; 990S-0144333.
PR 19 JUL-1999; 990S-0144334.
PR 19 JUL-1999; 990S-0144335.
PR 20 JUL-1999; 990S-0144335.
PR 20 JUL-1999; 990S-0144352.
PR 20 JUL-1999; 990S-0144632.
PR 20 JUL-1999; 990S-0144884.
PR 21 JUL-1999; 990S-0144814.
PR 21 JUL-1999; 990S-0145086.
PR 21 JUL-1999; 990S-0145088.
PR 22 JUL-1999; 990S-0145085.
PR 22 JUL-1999; 990S-0145087.
PR 22 JUL-1999; 990S-0145089.
PR 22 JUL-1999; 990S-0145192.
PR 23 JUL-1999; 990S-0145192.
PR 23 JUL-1999; 990S-0145145.
PR 23 JUL-1999; 990S-0145218.
PR 24 JUL-1999; 990S-0145224.
PR 26 JUL-1999; 990S-0145276.
PR 27 JUL-1999; 990S-0145913.
PR 27 JUL-1999; 990S-0145918.
PR 27 JUL-1999; 990S-0145919.
PR 28 JUL-1999; 990S-0145951.
PR 02 AUG-1999; 990S-0146386.
PR 02 AUG-1999; 990S-0146388.
PR 02 AUG-1999; 990S-0146489.
PR 03 AUG-1999; 990S-0147048.
PR 04 AUG-1999; 990S-0147204.
PR 04 AUG-1999; 990S-0147402.
PR 05 AUG-1999; 990S-0147192.
PR 06 AUG-1999; 990S-0147260.
PR 06 AUG-1999; 990S-0147403.
PR 06 AUG-1999; 990S-0147416.
PR 09 AUG-1999; 990S-0147494.
PR 09 AUG-1999; 990S-0147494.
PR 09 AUG-1999; 990S-0147945.
PR 10 AUG-1999; 990S-0148171.
PR 11 AUG-1999; 990S-0148419.
PR 12 AUG-1999; 990S-0148441.
PR 13 AUG-1999; 990S-0148565.
PR 13 AUG-1999; 990S-0148684.
PR 16 AUG-1999; 990S-0149468.
PR 17 AUG-1999; 990S-0149175.
PR 18 AUG-1999; 990S-0149426.
PR 20 AUG-1999; 990S-0149722.
PR 20 AUG-1999; 990S-0149723.
PR 20 AUG-1999; 990S-0149929.
PR 23 AUG-1999; 990S-0149902.
PR 23 AUG-1999; 990S-0149930.
PR 25 AUG-1999; 990S-0150566.
PR 26 AUG-1999; 990S-0150884.
PR 27 AUG-1999; 990S-0151065.
PR 27 AUG-1999; 990S-0151066.
PR 27 AUG-1999; 990S-0151080.
PR 30 AUG-1999; 990S-0151303.
PR 31 AUG-1999; 990S-0151488.
PR 01 SEP-1999; 990S-0151930.
PR 07 SEP-1999; 990S-0152363.
PR 10 SEP-1999; 990S-0154070.
PR 13 SEP-1999; 990S-0154758.
PR 15 SEP-1999; 990S-0154018.
PR 16 SEP-1999; 990S-0154039.
PR 20 SEP-1999; 990S-0154779.
PR 22 SEP-1999; 990S-0155139.
PR 23 SEP-1999; 990S-0155486.
PR 24 SEP-1999; 990S-0155659.
PR 28 SEP-1999; 990S-0156458.
PR 29 SEP-1999; 990S-0156596.
PR 04 OCT-1999; 990S-0157117.
PR 05 OCT-1999; 990S-0157753.
PR 06 OCT-1999; 990S-0157865.
PR 07 OCT-1999; 990S-0158029.
PR 08 OCT-1999; 990S-0158232.
PR 12 OCT-1999; 990S-0158369.
PR 13 OCT-1999; 990S-0159294.
PR 13 OCT-1999; 990S-0159294.
PR 13 OCT-1999; 990S-0159295.
PR 14 OCT-1999; 990S-0159429.
PR 14 OCT-1999; 990S-0159430.
PR 14 OCT-1999; 990S-0159441.
PR 14 OCT-1999; 990S-0159637.
PR 14 OCT-1999; 990S-0159648.
PR 18 OCT-1999; 990S-0159584.
PR 21 OCT-1999; 990S-0160741.
PR 21 OCT-1999; 990S-0160767.
PR 21 OCT-1999; 990S-0160768.
PR 21 OCT-1999; 990S-0160770.
PR 21 OCT-1999; 990S-0160814.
PR 21 OCT-1999; 990S-0160815.
PR 22 OCT-1999; 990S-0160980.
PR 22 OCT-1999; 990S-0160981.
PR 22 OCT-1999; 990S-0160989.
PR 25 OCT-1999; 990S-0161404.
PR 25 OCT-1999; 990S-0161405.
PR 25 OCT-1999; 990S-0161406.
PR 26 OCT-1999; 990S-0161359.
PR 26 OCT-1999; 990S-0161360.
PR 26 OCT-1999; 990S-0161361.
PR 28 OCT-1999; 990S-0161920.

PP 28-OCT-1999; 99US-0161992;
 PR 28-OCT-1999; 99US-0161993;
 PR 29-OCT-1999; 99US-0162142;

Query Match 33.7%, Score 825.5; DB 21, Length 493;
 Best Local Similarity 38.6%, Pred No 2 3e-67;
 Matches 180; Conservative 88; Mismatches 189; Indels 9; Gaps 2;

QY 20 GDAWID-----VVFATAVTSKIPDQAEIAAKKALUAARAKAUEWAEKALVAIGRWSW 71
 DB 23 GKWLDSDYNDKTIKVNPAIGELIADVACMGITKIDAIASIEAEFTSMKSLTAGERSKV 82
 QY 72 LRFISAGIRPASFSIALIVPEGKIQOLAEVEVAFADYIDVMAEWAPYEGETITQSDR 131
 DB 83 LRRWYDILLTAHKEELGQIILEGCKPLKEALGEVAGVAGSPFLEYAEAKVYSEIIPPNL 142
 QY 132 POENILLKRALGVTTGILPNWFFFLIARKMAPALLTGNTIVIKSEFTTNAIAFAKI 191
 DB 143 SDFRTIIVIKQFVGIVGATFENWFFLAMITRKVGIFALASCTVWVKTSILPIPLALAAEL 202
 QY 192 VDEIGLPRGVNLVGRGTGQELAGNPKVAMVMTGVSAGEKIMATAAKNITKVCLE 251
 DB 203 ALQAGVPGALNVWGNNAFELGALLTSPQVKIIFLGS:AVGKK:MAAANPVIKKVSLB 262
 QY 252 LOGKAPATIMDADLELAVKAIVDSRVINSQVNCNAERYVYVKGIVDQFVNPFGAMQA 311
 DB 263 LGSNAFSIVELGALLVAVKGILAAKFRNSJQTCVANKVLVGLGIVDAFAEAFSEAVQK 322
 QY 312 VQFGNPAERNDIAMPLINAAALERVQKVARAVEEGARVAFGKGVGYYPPTLLL 371
 DB 323 LEVGD-GRDGTGGLINDAAVQKVFVQDAVSKGAKIIGGKRHSGLGMYEPTVIR 381
 QY 372 DYRQMSIMHETFGVLPVVAFTDLEDAISMANDSDYGLTSSIIYUNLVAMKALGKLG 431
 DB 382 DVSNDKMSKEIEFGVAPLIRFKTEEDAIRANDTIAGLAAYFTNSVORSRVRFEALE 441
 QY 432 PCEITVINTEFMGGFIAGKWKSGICGACGCKHGHGCHYLOTQVYVL 477
 DB 442 YGLVAVNEGLISTEVAPEPRVKKQSGIGFPGSKYGMDEYLEIKVCL 487

RESULT 14

AB047807
 ID AB047807 standard; Protein: 488 AA

XX AC AB047807;
 XX DT 05-FEB-2002 (first entry)
 XX DE Listeria monocytogenes protein #511.
 XX KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
 XX KW vitamin B12; bacterial infection; disease.
 XX OS Listeria monocytogenes.
 XX PN WO2001177335-A2.
 XX PD 18-OCT-2001.
 XX PF 11-APP-2001; 2001WO-0001118
 XX PP 11-APP-2000; 2000EP-0004620.
 XX PA (INSP) INST PASTEUR.
 XX PI Ruchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
 PI Dussurget O, Chetolani F, Nedjari H, Glaser P, Kunst F, Cossart P;
 PI Daniels J, Goebel W, Kretz J, Kuhn M, Ng E, Vazquez-Boland JA;
 PI Dominguez-Bernal G, Garrido-Garcia F, Tierrez-Martinez A, Amend A;
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
 PI Madueno E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;

FI Rose M, Voss H,
 XX WF1, 2002-010914/01.

XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
 PT and prevention of Listeria and related bacterial infections, and
 PT related polypeptides

XX Claim 6; SEQ ID No 512; 192pp; French.

XX The present invention relates to the genome sequence of Listeria
 CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
 CC it are useful for selecting probes and primers for detecting genes in L.
 CC monocytogenes and related organisms, and for studying genetic
 CC polymorphisms and other genomes. The present sequence is a protein
 CC encoded by the genome sequence of the present invention. Proteins
 CC expressed from the genome sequence are useful for raising specific
 CC antibodies, identification of L. monocytogenes and related organisms, and
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
 CC B12. The genome sequence and proteins encoded by it are also useful for
 CC selecting compounds that regulate gene expression and cell replication
 CC and modulate L. monocytogenes-related diseases. In addition, the genome
 CC sequence and proteins encoded by it are useful in pharmaceutical and
 CC vaccines compositions for the treatment or prevention of infections by L.
 CC monocytogenes and related organisms.

XX Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 488 AA;

Query Match 33.2%, Score 811.5; DB 23; Length 488;

Best Local Similarity 46.4%; Pred No 4 4e-66;

Matches 175; Conservative 107; Mismatches 186; Indels 13; Gaps 4;

QY 5 VOHPMYIDQVTVRGDAWIDVNPATEAVLSRIPDQAEADAKKIDAARAPPEWALP 64
 DB 11 VQTKIFNGKWTGDKNETKIDVNPANVDVIAKJAGAPSETFKKAATAPDPAKMF 79
 QY 65 ATERPASWPKTSAGTRFPASRISALIVFEGGKIQOLAEVEVAFADYIDVMAEWAPYEFG 124
 DB 71 LADRVKLLKILADLMEEKADTLAKIMTLEQCKPLKESKGEVLGVENFPAEAEARLYG 130
 QY 125 EIIQSDPPGEN---IIFKPAIGVTTGILPNWFFFLIARKMAPALITGNTIVIKSEFT 181
 DB 131 ETI-----PAPNNHAFIVKKQPIGVVAAITPWNPPGCMVTRKIAPALATNTIVIKSEGDI 186
 QY 182 TNNATAFKIVDFEIGLPRGVENIVLGRGTGQELAGNPKVAMVMTGVSAGEKIMATA 241
 DB 187 PLSAIAIPEIFEEFASLGRGVANIVMGSSKKEIGETITSSIVAPKLEIFGRSKVQGITFKQS 246
 QY 242 AKNIITKVLELGSKKAIATVMEALALELAVKALVLSRVINSQVNCNAERYVYVKGIVDQF 401
 DB 247 ADTLKKISLELGGHAPFIVFDUANLDAAVNLVAAKFKNNQGVCSVSPNRTFVAKELKEF 406
 QY 302 VNPCEAMQAVGFONPAEPNDIAMSPLINAAALERVQKVARAVEEGARVAFGKGVGY- 460
 DB 307 TKALVAKVEQLKVGNGLD--DVNVVGPILIPFAIUKIUKQLKNAEKCAKYLITGRPELTQS 464
 QY 361 ---KGYVYPPPTLLDVPCEMSIMHETFGVLPVVAFTDLEDAISMANDSDYGLTSSIIY 417
 DB 365 DYDKGNFYKPTVLNDVTRKKMDIFYEETFGVPIPLITFETEDALEANSEGLASYFYI 424
 QY 418 QNLNVAMKALIKSLKRFETYINPNEFFAMQGFHAWRKSGSGGAAURKHGHLGCHYLOTQVYVL 477
 DB 425 KDLAPVERVGAALFVGMVGANEIAISNPPTPGCGVKHSGPGFENGHYGMFEYITQVRFINI 484
 QY 478 Q 478
 DB 485 K 485

RESULT 15

Search completed: June 24, 2003, 10:14:45
Job time : 40.6254 secs

QY 369 LLLDVROEIMSIMHEETGPGVPLVVAFDTLLEDAISMANDSDYGLTSSITONLNVAKAIFK 428
DB 383 VLDNRKDYVNFDEIFGCVLAITTYRDFEQVIEDANDTACGLSSVISENTEYMTATE 442
QY 429 GLKRGFYINRNFEAMQCGHAGWKSIGGAGLGRHSLHGLYGLQVWVLLQ 478
DB 443 RLKRGFYANCAEEVNGYHAGWRESGLGGADGIGHGFEYNTTVSYIR 492

RESULT 2

US-09-134-001C-4383
; Sequence 4383, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4383
; LENGTH: 506
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4383

Query Match 31.4% Score 767.5; DB 4; Length 506;
Best Local Similarity 36.6%; Pred. No. 1e-70;
Matches 171; Conservative 89; Mismatches 200; Indels 7; Gaps 3;

QY 10 YIDGQFVTHGDAWIDVVPATEAVISRIPTDGAEDAKAIDAAEPA--QPPWEALPAIE 67
DB 22 YIDGWEVSSNKNTRDIINPYNOETITFTVAETGKEDVERAILAARSRFEDGWSLETSEV 81
QY 68 RASWLKISAGIRERASEISALIVEREGGKIQQLAEVAFADYIDYMAEWARKYEGRII 127
DB 82 RGKVRVADKIKENRFEELAKLETIDTGKTLIESYADMDIHNVPWFYAGLAUKGSELI 141
QY 128 QSDRPGENILFLKRALGVITGILPNFPFELIARKMAPALLTGNTIVIKPSEFTNNIA 187
DB 142 NSIPNAPKSVKVPVGVVTOITPNWYPILOASWKIAPALATGCSLVMKPSEITPLTIR 201
QY 188 FAKIVDEIGLPGVFNVLGRGETVGOELAGNPKVAMYSMTGVSAGKIMATAAKNITK 247
DB 202 VPFLMEVETPKTINLVICATSEVGVMSGHEVDIVSFTGCTETGRHKWQANHVTD 261
QY 248 VCLHGLGKAPAVIMDDADLELAVKAIVDSPVINSQVNCAPVYVVKGIYDQFVNPIGE 307
DB 262 VALFLGKNPNIIFDDADFLAVDQALNGYFHAGQVCSAGSRILLVHNDIKDKFEKALID 321
QY 308 AMUAVQGNPAENDIAMPLINAAALEPVEKVPAPAFEGAPAPAFGKAVGK-----GY 363
DB 322 RVSKIKLGNFQO DTEMGVISTAHROKIEGYMEVAKKDGATIAIGKRPREDLQAGL 380
QY 364 YPPTLLDVROEIMSIMHEETGPGVPLVVAFDTLLEDAISMANDSDYGLTSSITONLNV 423
DB 381 FFEPTVTDGTSMRIVQVEEGPVVTVGSPADFEALPLANDSIYGLAGALFTKDICKA 440
QY 424 MKAIGLKGCEIYINRNFEAMQCGHAGWKSIGGAGLGRHSLHGLYGLQVWVLLQ 470
DB 441 QRVANKIKLGTVMINDFHPFAQAPGWGKYQSGIGRELCKEGLLEYL 487

RESULT 3

US-09-134-001C-4246
; Sequence 4246, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4246
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4246

Query Match 30.2% Score 737.5; DB 4; Length 488;
Best Local Similarity 35.8%; Pred. No. 1.2e-67;
Matches 169; Conservative 93; Mismatches 195; Indels 15; Gaps 7;

QY 10 YIDGQFVTHGDAWIDVVPATEAVISRIPTDGAEDAKAIDAAEPA--QPPWEALPAIE 69
DB 21 YINGPWWDSASGRTTIVINPATEFVVMCKTAGCNFTFVNPAPVPAACVYVTFTHSSVEERR 80
QY 70 SWLEKISAGIRERASEISALIVEREGKIQQLAEVAFADYIDYMAEWARKYEGRII 126
DB 81 ELIDKIVKRYONKKNULIKATIDELG AULSVSENVHYQMGHIFT AAKGALIS 144
QY 127 IQSDRPGENILFLKRALGVITGILPNFPFELIARKMAPALLTGNTIVIKPSEFTNNIA 186
DB 134 FQEEQGGNULVVKKEAIGVAGSLVTPWNPFTNUTSLKLAFAAAGSPVVKPSEFTPEAA 194
QY 187 AFAKIVDEIGLPGVFNVLGRGETVGOELAGNPKVAMYSMTGVSAGKIMATAAKNIT 246
DB 194 ILAEIFKGVKVPVGVNVLNGDSGVNLSLHPVPMMSFTGSPSTSEINERKAATPK 254
QY 247 KVCFLGSKAPAVIMDDADLELAVKAIVDSPVINSQVNCAPVYVVKGIYDQFVNPIGE 306
DB 254 KVSLELGKSPYIVLDSDVVEEANAATTKKKVNNITQCVTACTRVLIPESIKEDYLTAVK 414
QY 307 EAMUAVQGNPAENDIAMPLINAAALEPVEKVPAPAFEGAPAPAFGKAVGK VEG KAY 464
DB 314 EAFSKYKVGQPPREG TQWPTLSKKQFQVGVPIIRKGINEGAELEYGGGKRPGLDGGY 472
QY 364 YPPTLLDVROEIMSIMHEETGPGVPLVVAFDTLLEDAISMANDSDYGLTSSITONLNV 424
DB 373 FAPRTIFINVDNHTITAOEILFCPVMSVITYNNILDEAIEIANDKYGLAGYVICKIKITL 442
QY 424 MKAIGLKGCEIYINRNFEAMQCGHAGWKSIGGAGLGRHSLHGLYGLQVWVLLQ 470
DB 443 PHVAPSTEATTFINPAEPKPLPE GYKESGIRGFWGDFEELFKVSI 488

RESULT 4

US-09-513-841-2
; Sequence 2, Application US/08513841
; Patent No. 5753481
; GENERAL INFORMATION:

; APPLICANT: Niwa, Mineo
; APPLICANT: Saito, Yoshinori
; APPLICANT: Ishii, Yoshinori
; APPLICANT: Yoshida, Masaru
; APPLICANT: Suzuki, Hiromi
; TITLE OF INVENTION: No. 5753481el L-sorbitose dehydrogenase and No. 5753481el L-s
; NUMBER OF INVENTIONS: dehydrogenase obtained from Gluconobacter oxydans T 100
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: obion, Spivak, McJelland, Maier & Neustadt, P.C.
; STREET: 1755 Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA

CORRESPONDENCE ADDRESS:

ADDRESSEE: Orlon, Spivak, McClelland, Maier & Neustadt, P.C.
 STREET: 1755 Jefferson Davis Highway, Suite 400
 CITY: Arlington
 STATE: Virginia
 COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1 44 Mb storage
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: MS-DOS Editor
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/118,317
 FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/513,841
 FILING DATE: 01-NOV-1995
 APPLICATION NUMBER: UK 9304700 9
 FILING DATE: 08-MAR-1993

PRIOR APPLICATION DATA: JP 241851/1993

FILING DATE: 28 SEP-1993

ATTORNEY/AGENT INFORMATION:

NAME: NORMAN F. OBLON

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 18-909-0 PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-413-3000

TELEFAX: 703-413-2220

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO. 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 497 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

ORIGINAL SOURCE:

ORGANISM: Gluconobacter oxydans

STRAIN: T-100

FEATURE:

NAME/KEY: mat peptide

LOCATION: 1..497

IDENTIFICATION METHOD: experimentally

US-09-118-317-2

Query Match 27 48, Score: 683.5, DB 4, Length: 497.

Best Local Similarity 33.18, Field No. 50-62.

Matches 159; Conservative 107; Mismatches 206; Indels 9; Gaps 6;

QY 4 PVQHPMYIDGQFVWKGAWIIIVVNPATEAVISRIPIQGLAEIARAKAIDAAEKA--QPEW 60
 DB 13 PREGFFIDGE---WRAGKDFDFDSSPAHDVPTPIPTCTPDLEAVAAAPAPENSW 69
 QY 61 EALPAIEPASMIRKTSNGTPERASPTISALIVERGKIQQLAEVEVAFTADYIDYMAEWA 120
 DB 70 AGIAAADRAAVIILAKAGLLRRERDIAYWEVLNENKPTISQAKGEIDHCITACPEMAAGA 129
 QY 121 RYGEITQSDRRPCEHILFKPAIGVITIGILPNNPFEELIAPKMAPALLTGNITIVIKPSEF 180
 DB 130 MLHGDTNNLGEGLFGMLVREPIGVVGLITPNNPPEMLCECAPFTLASGCTLVVYKPAEV 189
 QY 181 TTNNIAFAKIVIDEISLPGFVNLVLRGETVQGLAGNPKVAMVSMIGSVSAGEK--IMA 239
 DB 190 TSAITLLLAELADAGLPKGVENVVGTGRTVQGMTEHGD:DLMSFTGTCGVGKSCIIIA 249
 QY 240 TAANKITKVLGIAGKAPATVMPDADLEIAKVAIVTSPVINSQVNCNCAERYVYVKGIVD 299
 DB 250 AADSNLKKIGLIGLGGKNP1VVFALSNLELAAGAVAFGLSFN:QQCCVSSSRILIVERKVAE 309
 QY 300 QFVNPIGEAMCAVCFNPAPKNDIAM3PLINAAALEPVEUKYAPAVFPAKAVFAGKAVE 359

Lb 310 KFEELVVVFMKELFVSDFFS--FETQLCAIIIEA2NPKIILLYIAK:KAE:AKVLLDGGIVD 368
 QY 369 GR3YVYPTLILLOVPEMSIMHEFTFGVILVVAHGLLELALSMANESYGLISSIVTQ 418
 DB 369 FGRGQYLCPTLFLHVRKPSMGIAHPTFGFVLASPHPTVIEAIAIANDTVY:LAASVWSK 428
 QY 419 NLNVAMKAIKSLKEGETIYNFENFEAMQGFHAGWKSGLGGAGDKHGLUGLYLOTQVVVILQ 478
 DB 429 ITHKALAVTPKVPAGPFWNTIMSGPPTF4:3FKSGSWCPRAGLYGVVEYTOIKSVHIE 488
 QY 479 S 479
 DB 489 T 489

RESULT 8

US-09-221-294-2

Sequence 2, Application US/09221294

Patent No. 6268138

GENERAL INFORMATION:

APPLICANT: Riccardo Dalla-Favera and

APPLICANT: Alessandro Massimo Gianni

TITLE OF INVENTION: A Putt-vital V-v-v-v-v Capable of Transducing the

TITLE OF INVENTION: Aldehyde Dehydrogenase-1 Gene and Uses of Said

TITLE OF INVENTION: Vector

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSER: Cooper & Dunham LLP

STREET: 1185 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM 330 466 DX2

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.24

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/221,294

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 42990-B

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-278-0400

TELEFAX: 212-391-0525

INFORMATION FOR SEQ ID NO. 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 521 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-221-294-2

Query Match 27.48; Score 669.5; DB 4; Length 521;

Best Local Similarity 34.44; Field No. 150-60;

Matches 159; Conservative 89; Mismatches 197; Indels 17; Gaps 9;

QY 26 VWFATFAVTSPIHMGAGAHAPKAI:AAIQA QFWALCA:AFASWIPKISASIPER 82
 DB 56 VFNPAEELCQVEFGTKFTVDKAVKAAQAFQTSQSPWPTMPASERGLLYKL-ADLIEP 114
 QY 83 ASPLSALIVE---GQKILCLAFV-KVAFTAVYIDYMAKWAFFYEGE--ILQSDRGENT 136
 DB 115 DRLLATMESNMNGKLYSNAYINDLAGCIKTRFCAGWADKIOGGKTFID---GNF 171
 QY 137 ILFKP--ALGVGTILPWNFFELIAFKMAPALLTGNITIVIKPSEFTNNIAIAFAKIVDE 194
 DB 172 FTYTRHEPICVCGGITWNTFTVLM:WKIGFALSCGNTVVVKFAEGTFLIALHVASLIKE 231
 QY 195 LSLPPGVFNVLVSGPGRIVGJELAGNPKVAMVSMIGSVSAGEK--IMAIAAK-NITKV:LELGS 253

```

Db 232 AGPPGVNVPVGYGTAGAAISSHMDIDKVAFTSTGVGKLIKAAAGSKNLKRTVLLEIG 291
Qy 254 GKAPAVMDADLELAVKALVDSRVINSQVCNCAERYVVKGIYDQFVNRLGEMAOAVQ 313
Db 292 GKSPCLVLAADLNAVEFAHGVFYHOGQCCIAASRI FVERIS IYDEFVRKSRKYYI 351
Qy 314 FGNPAENDIAMGPIINAAALERVOKVARAVEEGARVAFGGKAVGGKGYPPPTLLLDV 373
Db 352 LGNPLTPG-VTQGPQDKQYQDKITDGFSGKKGKAKLFGGCGPWCNGKGYEVQPTFSNV 410
Qy 374 RQMSIMHETFGVLPVVAFLDLELALSMANUSYGLTSSLYIYNLNVAMKAIKGLKFG 433
Db 411 TDMRIAKERIFGVVQQIMKPKSLDIVIKRANNTFYGLSGAGVFTKIDKAITISSALQAG 470
Qy 434 ETVINKENFAMGGFIAGWKSGIGGAGCKHGLHGYLQTVV 475
Db 471 TVWVNCYGVVSAQCPRGFGKMSONGRELGEYGFHEYTEVKT 512

RESULT 9
US-09-155-183-4
; Sequence 4, Application US/09155183
; Patent No. 6323011
; GENERAL INFORMATION:
; APPLICANT: Narbad, Arjan
; APPLICANT: Rhodes, Michael J.C.
; APPLICANT: Gasson, Michael J.
; APPLICANT: Walton, Nicholas J.
; TITLE OF INVENTION: PRODUCTION OF VANILLIN
; FILE REFERENCE: 20747/100
; CURRENT APPLICATION NUMBER: US/09/155,183
; EARLIER FILING DATE: 1999-05-03
; EARLIER APPLICATION NUMBER: PCT/GB97/00809
; EARLIER FILING DATE: 1997-03-24
; EARLIER APPLICATION NUMBER: GB96/06187
; EARLIER FILING DATE: 1996-03-23
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 4
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Pseudomonas fluorescens
US-09-155-183-4

```

```

Query Match 26.1%; Score 630.5; DB 4; Length 482.
Best Local Similarity 33.1%; Pred. No. 1.7e-57;
Matches 159; Conservative 88; Mismatches 216; Indels 17; Gaps 7;

```

```

Qy 8 PMYIDGQFVTWRCTAMITVVNPATEAVTSRPTGSAFLAPKATUAATFAQFWKALIAIF 47
Db 5 PLLIGQSCFARKGRTPFRPNPTGTGLSPVAAAATLEDAADAAVAAQAQAFPAWALAPNE 64
Qy 68 RASWLRKISAGTIPERASRISALITVEGCKIQQIAFA---VEVAFATGYIDYMAFWAPPEG 124
Db 65 PRSPTIKAAQVQAQSCG-----FIFAAGETGAMAWCYFNVLKLANMLREASMTTVNCC 129
Qy 125 EITQSDRPGENTLLKRALGVTTGILPWNPFPFLIARKMAPALITGNITVIKPSFEFTINN 184
Db 121 EVIPLDVPFGSAMALRQPGVVLGIAPWNPVILATRAIAMPLACGNTVLKASELSPAV 180
Qy 185 AIAFAKIVDFEIGLPKCVFNILVLCRGETVQ---ELAGNPKVAMVSMTCGYSAGCKINATA 241
Db 181 HPIIGQVLAQAGIAGGVVNVNINAPADAQAQIVERTIANPVPVPPNPGTSHVPIVIGELS 249
Qy 242 AKHITKVCLEGGKAFAPIMIDADI ELAVKALVESFVINSQVCNCAERYVVKGIYDQFVNRL 301
Db 241 ARHLKPAILELGGKAPILLVDADLEAAVQAAAFCAYNOCISTERTLIVDAKVAADAP 300
Qy 302 VNLKLGEMAOVQFGNPAENDIAMGPIINAAALERVOKVARAVEEGARVAFGGKAVGGK 361
Db 301 VAGIAAKVETIRAGDPAHPESV-LGSLVDASAGTRIKALIDCAVAKGARILVIGGQL---E 356

```

```

Qy 362 GYYPTPTLLLDVQKQMSIMHETFGVLPVVAFLDLELALSMANUSYGLTSSLYIYNL 421
Db 357 GSTLOPTLLDGVASMBLYRSPSPGVAVVLGEGEALQIANISERHSAAITPSRDIG 416
Qy 422 VAMKAIKIKFKEFTYINFEF FAMQCPHAWRKSGTICALCKRCHRYVLTQTVVYIQS 479
Db 417 KALALAQRVESGLCHINGTIVHEAUMPE-GVKSNGYSGPGKASTHEFTQLRWVILON 475

RESULT 10
US-09-134-001C-4541
; Sequence 4541, Application US/09144001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: CTC-007
; CURRENT APPLICATION NUMBER: US/09/144,001C
; CURRENT FILING DATE: 1998-08-14
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4541
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4541

```

```

Query Match 25.9%; Score 633.5; DB 4; Length 510;
Best Local Similarity 30.6%; Pred. No. 4e-57;
Matches 146; Conservative 100; Mismatches 222; Indels 9; Gaps 6;

Qy 9 MYTIDGQFVTWRCTAMITVVNPATEAVTSRPTGSAFLAPKATUAATFAQFWKALIAIF 68
Db 29 LFINPEQASDSGETLTVSNIPANGDELAKVAPAKKIKVIRKAVQAARLAFISWSKISKER 88
Qy 69 ASWIKPKISAGTIPERASRISALITVEGCKIQQIAFA---VEVAFATGYIDYMAFWAPPEG 127
Db 89 ADYLETSSKRTKERTHEILAVESLQNGKPYRRTSTLTVQAANQKYFASVLTIDRESVN 148
Qy 128 QSDRPGENTLLKRALGVTTGILPWNPFPFLIARKMAPALITGNITVIKPSFEFTINNATA 187
Db 149 FTLQNTMSLAVVNFV-VVGTVVAMRPTLTAANKLGLALAA-INAVV-QVTSSTVLSLE 209
Qy 188 FAKLVDEIGLPKCVFNILVLCRGETVQ---ELAGNPKVAMVSMTCGYSAGCKINATA 247
Db 208 LAKITGVVTPKGVVNVNINAPADAQAQIVERTIANPVPVPPNPGTSHVPIVIGELS 266
Qy 248 VTELEGGKAFAPIMIDADI ELAVKALVESFVINSQVCNCAERYVVKGIYDQFVNRL 307
Db 267 TTELEGGKAFAPIMIDADI ELAVKALVESFVINSQVCNCAERYVVKGIYDQFVNRL 326
Qy 308 AMCAVQCFMIALNRIAMGPIINAAALERVOKVARAVEEGARVAFGGKAVGGK 362
Db 327 AFENKVRQVDEI-EITKNSAGTIGELQIFESYKRIAEETKANEITGAPHTLNNGLK 385
Qy 363 YYYPTPTLLLDVQKQMSIMHETFGVLPVVAFLDLELALSMANUSYGLTSSLYIYNL 422
Db 386 YFFPTPTLLLDVQKQMSIMHETFGVLPVVAFLDLELALSMANUSYGLTSSLYIYNL 445
Qy 423 AMFAIKIKFKEFTYINFEF FAMQCPHAWRKSGTICALCKRCHRYVLTQTVVYIQS 479
Db 446 ALNVAKAMFISFVWLNINSGTIFAAAPFGKSGFSGTIPVYVAFAPNGLVNFIDP 502

RESULT 11
US-09-134-001C-4451
; Sequence 4451, Application US/09144001C
; Patent No. 6380370
; GENERAL INFORMATION:

```



```

QY 127 IQSDRPGENI-LLFKRALGVTTGILPWNPPFLLIARKMAPALLTGNTIVIKPSEFTTNA 185
DB 128 -TTPAPGRFNYTLRPIGVGGAITWNEFPAVQVWKIAPALAMGNAIVLKPALVLP 186
QY 186 IAFKIVDEIGLPRGVNVLGRGTGVLGAGNPKVAMVMTGVSAGEKIMATAAKNI 245
DB 187 VALFETALFASLPPGLVNVTPRPSVAQNALVQHSVQKVIETSEFVAGLQIPMAACPVL 246
QY 246 TKVTEIGRKAIPALVMLQALIELAVKRAIVDSVINSQW-N-AHVVYVUKGILYQFVNL 305
DB 247 ITASLEIGGKLSALVAFGDSKPAVAAVFQAMYSNOGETCTAPSRLLVERPIYDEVVELV 306
QY 306 GEAMQAVQFQNPAPERNTIANGPLINAAALFRVEQKVAPEEGARVAFGG----KAVEGK 361
DB 307 QARVEARVGDPLD-PDTEIGPLISAEQRESVHSVVSSTEGEATLISGGDSPTGAPEDQ 365
QY 362 GYYPPTLLIDVPQMSIMHEETFGVLPVVAFTLEDAISMANSDYGLTSSITYONIN 421
DB 366 GFYRPTLFGVTDAMRIAREFTFGPVLVLPPEGEETATLANTVFLAAGVFTPDVG 425
QY 422 VAMKATKLGFTGYINENFAMQGFHAGWKEKSLGGALGKHGLGVLQTVVY 476
DB 426 RALRFAOTLDAGNWNINSWCVINPASPYPGQSGYSDLGQAATESFTKEKSIW 480

RESULT 14
US-09-655-270A-9
; Sequence 9, Application US/09655270A
; Patent No. 6329151
; GENERAL INFORMATION:
; APPLICANT: Rouviere, Pierre E.
; TITLE OF INVENTION: High Density Sampling of Differentially Expressed Prokaryotic m
; FILE REFERENCE: H01011 US NA
; CURRENT APPLICATION NUMBER: US/09/655,270A
; CURRENT FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/120,702
; PRIOR FILING DATE: 1999-February-19
; PRIOR APPLICATION NUMBER: 60/152,542
; PRIOR FILING DATE: 1999-September-03
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 9
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Rhodococcus erythropolis HL PM-1
US-09-655-270A-9

```

```

Query Match 35.5%; Score 623.5; DB 4; Length 508.
Best Local Similarity 29.9%; Pred. No. 8.6e-56;
Matches 142; Conservative 99; Mismatches 225; Indels 9; Gaps 5;

QY 8 PVIYDQGVTVWRGDAMIDVWNPATEAVISKIPUGAELAKKALIAAARAUPWEALPAIE 67
DB 32 PLVIGLQITPSSGATEDSINPADGSHLASVAETAADVARAVEAKAAARTWORMRFAQ 91
QY 68 RASWLKLSAGITRERASEISALIVEGK-IQGLAEVEVAFYADYIDYMAEWARRYECEI 126
DB 92 RTRLMFRYALTEEHKTELAQOSRDMCKPIRESIGDLPIMETLFYFAGLVTKIEGR- 150
QY 127 IQSDRPGENI-LLFKRALGVTTGILPWNPPFLLIARKMAPALLTGNTIVIKPSEFTTNA 185
DB 151 -TTPAPGRFNYTLRPIGVGGAITWNEFPAVQVWKIAPALAMGNAIVLKPALVLP 209
QY 186 IAFKIVDEIGLPRGVNVLGRGTGVLGAGNPKVAMVMTGVSAGEKIMATAAKNI 245
DB 210 VALGELALEAGLPPGLVNVTPRPSVAQNALVQHSVQKVIETSEFVAGLQIPMAACPVL 269
QY 246 TKVTEIGRKAIPALVMDADLELAVKRAIVDSVINSQWNCARVYVUKGILYQFVNL 305
DB 270 ITASLEIGGKLSALVAFGDSKPAVAAVFQAMYSNOGETCTAPSRLLVERPIYDEVVELV 329
QY 306 GEAMQAVQFQNPAPERNTIANGPLINAAALFRVEQKVAPEEGARVAFGG----KAVEGK 361

```

```

DB 330 QARVEARVGDPLD-PDTEIGPLISAEQRESVHSVVSSTEGEATLISGGDSPTGAPEDQ 388
QY 362 GYYPPTLLIDVPQMSIMHEETFGVLPVVAFTLEDAISMANSDYGLTSSITYONIN 421
DB 389 GFYRPTLFGVTDAMRIAREFTFGPVLVLPPEGEETATLANTVFLAAGVFTPDVG 448
QY 422 VAMKATKLGFTGYINENFAMQGFHAGWKEKSLGGALGKHGLGVLQTVVY 476
DB 449 RALRFAOTLDAGNWNINSWCVINPASPYPGQSGYSDLGQAATESFTKEKSIW 508

RESULT 15
US-09-351-224E-5
; Sequence 5, Application US/09351224E
; Patent No. 6388171
; GENERAL INFORMATION:
; APPLICANT: Dupick, Jon
; APPLICANT: Maddox, Joyce
; APPLICANT: Gilliam, Jacob
; APPLICANT: Folkerts, Otto
; APPLICANT: Grastia, Oswald K.
; TITLE OF INVENTION: Compositions and Methods for Eumenisin
; FILE REFERENCE: Detoxification
; CURRENT APPLICATION NUMBER: US/09/351,224E
; CURRENT FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Exophiala spinifera
US-09-351-224E-5

```

```

Query Match 18.3%; Score 448; DB 4; Length 487.
Best Local Similarity 27.9%; Pred. No. 1.3e-47;
Matches 135; Conservative 85; Mismatches 212; Indels 12; Gaps 14;

QY 9 MYIDQGVTVWRGDAMIDVWNPATEAVISKIPUGAELAKKALIAAARAUPWEALPA 65
DB 12 LFINNEFVSSKSGERLILINPWUESIVALDIWIVANAADVSAVASVUAKKGWKKFTG 71
QY 66 ITHASWLKLSAGITRERASEISALIVEGK-IQGLAEVEVAFYADYIDYMAEWARRYECEI 124
DB 72 AQPAACMLKFAADLAEKNAEKLEPESLTGPRVSMITHETIPNMVSVFERYVAMADKIA 131
QY 125 EITIOSRPGENILFKRALGVTTGILPWNPPFLLIARKMAPALLTGNTIVIKPSEFTTNN 184
DB 132 KTEPELNGKNNRY--EPMGVVAGIASWNAIFLYWVKIALAALAAQSFTRKASEKSLG 189
QY 186 ALAPATVDEIGLPRGVNVLGRGTGVLGAGNPKVAMVMTGVSAGEKIMATAAKNI 243
DB 190 VLGALALFAAGLPPGLVNVTPRPSVAQNALVQHSVQKVIETSEFVAGLQIPMAACPVL 248
QY 244 NTKVTEIGRKAIPALVMDADLELAVKRAIVDSVINSQWNCARVYVUKGILYQFV 302
DB 249 NMKPVTLDEGK-PTIVNEAPLEPESGESAPF--FSPFPLTWVHPS-LLVWV 309
QY 303 NREGAMQAVQFQ-----NVARNDIAMGPLINAAALFRVEQKVAPEEGARVAFGG 361
DB 300 -NLAEPKFCVPRGSGQVPMIQQNDLPKQ-THGPPVQKSGYQDVIQNTIIVKRTIQAQL 366
QY 352 AFGKAVEGKGYVYVPTLLILVWQMSIMHEETFGVLPVVAFTLEDAISMANSDYGL 411
DB 357 LTGVGKKGKQKPALEPTFVNPDKSGKLVFEETFGPVLSTKTERTEETATLANTV 416
QY 412 TSSITYONINVAMKALKGLKFTGYINENFAMQGFHAGWKEKSLGGALGKHGLGVLQ 471
DB 417 ASVITKSLNGPLGVSSALETSVNSITPFTCTTPPGRKQSLSPELGEELKATLE 476
QY 472 TQVV 475

```

Tue Jun 24 10:42:22 2003

us-09-830-751-6.rai

Page 9

Db 477 PKT1 480

Search completed: June 24, 2003, 10:31:42
Job time : 15.2654 secs

GenCore version 5.1.6
Copyright (c) 1994 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:17:21 : Search time 23.1541 seconds
(without alignments)
2238 528 Million cell updates/sec

Title: US-09-830-751-6
Perfect score: 2446
Sequence: 1 MSVPVQHPMVIISQFVTRGALSKKHGLHAYLGLGVVYLS 479

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched 41772 seqs, 16826813 residues 41779

Total number of hits satisfying chosen parameters

Minimum DB seq length: 6
Maximum DB seq length: 200000000

Post-processing: Minimum Match 8%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA:

- 1: /cgn2_6/prodata/2/pubaa/us08_NEW_PUB pep.*
- 2: /cgn2_6/prodata/2/pubaa/PCT_NEW_PUB pep.*
- 3: /cgn2_6/prodata/2/pubaa/us06_NEW_PUB pep.*
- 4: /cgn2_6/prodata/2/pubaa/us06_PUB pep.*
- 5: /cgn2_6/prodata/2/pubaa/us07_NEW_PUB pep.*
- 6: /cgn2_6/prodata/2/pubaa/us07_PUB pep.*
- 7: /cgn2_6/prodata/2/pubaa/PCTUS_PUB pep.*
- 8: /cgn2_6/prodata/2/pubaa/us08_PUB pep.*
- 9: /cgn2_6/prodata/2/pubaa/us09_NEW_PUB pep.*
- 10: /cgn2_6/prodata/2/pubaa/us09_PUB pep.*
- 11: /cgn2_6/prodata/2/pubaa/us10_NEW_PUB pep.*
- 12: /cgn2_6/prodata/2/pubaa/us10_PUB pep.*
- 13: /cgn2_6/prodata/2/pubaa/us60_NEW_PUB pep.*
- 14: /cgn2_6/prodata/2/pubaa/us60_PUB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Query	Score	Match	Length	DB	ID	Description
1	879.5	36.0	483	10	US-09-815-242-5058		Sequence 5058, Ap
2	873.5	35.7	482	10	US-09-815-242-10264		Sequence 10264, A
3	858.5	35.1	482	10	US-09-815-242-14047		Sequence 14047, A
4	830	33.9	492	9	US-10-268-518-4		Sequence 4, Appl
5	819.5	33.5	493	9	US-10-175-696-21		Sequence 21, Appl
6	819.5	33.5	493	10	US-09-823-901-9		Sequence 9, Appl
7	771.5	31.5	496	10	US-09-815-242-5644		Sequence 5644, Ap
8	771.5	31.5	496	10	US-09-815-242-12657		Sequence 12657, A
9	768.5	31.4	481	10	US-09-815-242-13829		Sequence 13829, A
10	762	31.2	496	10	US-09-815-242-19550		Sequence 10550, A
11	756.5	30.9	475	10	US-09-815-242-12375		Sequence 12375, A
12	748.5	30.5	501	9	US-09-444-882-20		Sequence 20, Appl
13	748.5	30.5	501	9	US-10-293-865-22		Sequence 22, Appl
14	729	29.8	490	9	US-09-738-626-6572		Sequence 6572, Ap
15	729	29.8	490	10	US-09-815-242-12102		Sequence 12102, A
16	719.5	29.4	512	9	US-09-971-403-12		Sequence 12, Appl
17	719.5	29.4	512	9	US-10-268-518-2		Sequence 2, Appl
18	719.5	29.4	512	9	US-10-205-823-14		Sequence 14, Appl
19	712	29.1	490	10	US-09-815-242-10057		Sequence 10057, A

20	712	29.1	508	1	US-08-781-986A-5241	Sequence 5241, Ap
21	710	29.0	518	9	US-09-919-039-143	Sequence 143, Appl
22	706.5	28.9	534	9	US-09-444-882-24	Sequence 24, Appl
23	706.5	28.9	534	9	US-10-203-865-24	Sequence 24, Appl
24	700.5	28.6	538	9	US-09-444-882-20	Sequence 20, Appl
25	700.5	28.6	538	9	US-10-293-865-20	Sequence 20, Appl
26	691.5	28.3	487	9	US-10-175-696-17	Sequence 17, Appl
27	691.5	28.3	487	10	US-09-823-901-5	Sequence 5, Appl
28	681	27.8	496	9	US-09-847-208-59	Sequence 59, Appl
29	670	27.4	500	9	US-10-166-087-4	Sequence 4, Appl
30	662.5	27.1	498	9	US-10-272-419-12	Sequence 12, Appl
31	657.5	26.9	495	9	US-09-847-208-11	Sequence 11, Appl
32	643.5	26.3	491	10	US-09-997-664-3	Sequence 3, Appl
33	643.5	26.3	491	10	US-09-738-626-4	Sequence 4, Appl
34	636	26.0	504	9	US-09-738-626-6572	Sequence 6572, Ap
35	625	25.6	484	9	US-09-748-626-6430	Sequence 6430, Ap
36	610	24.9	454	9	US-09-748-626-4037	Sequence 4037, Ap
37	597	24.4	481	1	US-08-978-306-28	Sequence 28, Appl
38	592.5	24.3	434	10	US-09-815-242-13516	Sequence 13516, A
39	578.5	23.7	407	9	US-09-738-626-4108	Sequence 4108, Ap
40	529	21.5	521	9	US-09-748-626-6482	Sequence 6482, Ap
41	498	20.4	314	10	US-09-925-300-1348	Sequence 1348, Ap
42	484.5	19.6	504	9	US-09-748-626-6680	Sequence 6680, Ap
43	457.5	18.7	412	10	US-09-925-302-736	Sequence 736, Appl
44	448	18.3	487	9	US-09-847-694-5	Sequence 5, Appl
45	445.5	18.2	492	10	US-09-741-694-405	Sequence 406, Appl

ALIGNMENTS

RESULT 1
US-09-815-242-5058
: Sequence: 5058, Application US/09/815242
: Patent NO. US20020061369A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Kari L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Hawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: FILE REFERENCE: ELITRA 011A
: CURRENT APPLICATION NUMBER: US/09/815.242
: PRIORITY FILING DATE: 2001-03-21
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/243,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/259,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 5058
: LENGTH: 483
: TYPE: PRT
: ORGANISM: Pseudomonas aeruginosa
US 09 815 242 5058
Quality Match 36.04, Score 879.5, DB 10, Length 483.
Best Local Similarity 29.39, Pctd N. 2 39.68;

SEQ ID NO 14047
LENGTH: 482
TYPE: PRT
ORGANISM: Salmonella typhi
US-09-815-242-14047

Query Match 35.1%; Score 858.5; DB 10; Length 482;
Best Local Similarity 39.2%; Pred. No. 1.5e-66;
Matches 181; Conservative 91; Mismatches 189; Indels 1; Gaps 1;
QY 10 YIDGQVTWRGDADIVVNPATFAVIRTPINGQAEAPKATDAEAAPQWEALPALEPA 69
DB 14 FIDGQVTPAPAGGVTPVSNPANGKPTGNVPMKAEETPDATNAANPALPAMPALTAKEPA 73
QY 70 SWLKKISALINERKASEISALIVERGSKLQIARVEVAFTADYIDYAEWARRVEGELIOS 129
DB 74 NLRWFNIMMEHODDILARMTLEGGKPLAFAGKEISYAASTFWEAEGRKRYGDTIPG 133
QY 130 ERFQENILFKKALAVTTRGILFWNFPFLIARKMAHALLAGNTIVIKFSEFTTNNALAPA 189
DB 134 HQIDKRLIVIKOPLGVTAATITPWFNPSAMITRKAGPALAAGCTMWLKPASOTPFSALEPA 193
QY 190 KIVDEIGLPPGVNVLVGPQTVQQLAGNPKVAMVMTGVSAGKIMATAAKNTKVT 249
DB 194 ELAQAGAGIAGVFNVTGSDAGIGELTSNPVPLKLSFTGSTEIGFQLMQCCAKDKKVS 253
QY 250 IELGCKAPAIYMDADIEFLAVKATVDSRVNSGGVNCAPVYVVGKICVDOFVNPLGEAM 309
DB 254 LEIGGNAPFIVDDADIDKAVGALASKERNAGQTCVCANRLYVQGVYDRFAEKUNQAV 313
QY 310 QAVQGNPAERNDIAMGLINAAALERVEQKVARAVEGAVFAGKAVGKGYYPPTL 369
DB 314 NKLAVSDGLQA-DVAIGLIDFKAVAKVQPHIAQALEKGARVITGGSEAHKLGNGFTQPTI 372
QY 370 ILDPVGEKSMHETFGVPLVPAFTLELAISMANSDGLTSSIVTQNLNVMKAIKG 429
DB 373 LADVPDNKAKVEETFGPLAFRFSDEADVIPQANITEFGLAAIFYAPDI-SVPFPVGEA 432
QY 430 LKFGETYINRENFEAMGFHAGWRKSGTGAGDKKHGLHGYIQ 471
DB 433 LEYGVIGINTGISNEVAPFGIKASGLREGSKYGIEDYLE 474

RESULT 4

US-10-268-518-4
Sequence 4, Application US10268518
Publication No. US20030100034A1
GENERAL INFORMATION:
APPLICANT: Hunter, John Joseph
TITLE OF INVENTION: A HUMAN ALLELYNE (PHYIIP) GENES
FILE REFERENCE: MPI01-234P1RM
CURRENT APPLICATION NUMBER: US10/268,518
PRIOR FILING DATE: 2002-10-10
PRIOR APPLICATION NUMBER: 60/424,844
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 492
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Consensus sequence
US-10-268-518-4

Query Match 33.9%; Score 830; DB 9; Length 492;
Best Local Similarity 41.4%; Pred. No. 4.8e-64;
Matches 203; Conservative 78; Mismatches 175; Indels 34; Gaps 14;
QY 23 WFD-----VWNPATEA-VISRIPTGGAEPAPKATGAAPFA---QFWEALPALEPAS 79
DB 2 WDSAGSKTFEWNPNKKEVIGVPEATAEDVDAVKAKEAFKSGPWWAKVPASERAR 61

QY 71 WPKTSAGIPRPAISEISALIVERGSK--TQGLARVEVAFTADYIDYAEWARRVEGELIOS 126
DB 62 TLPLKADLJFRERDELAALFTLDLQKPLAEAKGCTEVGRAIDIRYVAGMARKIMGERRV 121
QY 127 IQS-DRPGENTILLEKR--ALGVTTGILPWNFPFLIARKMAPALLTGNITIVIKFSEFTTN 183
DB 122 IFSLATDDEELNTRKREPLGVGVVISFNPFPPLLLALWKLAPALAAAGNTVWLKPSQTP 181
QY 184 NALAPAKIVDRIG---LPRGVNVLVGRGETVQQLAGNPKVAMVMTGVSAGKIM-A 239
DB 182 TALLAELELEAGANNLPKVVNVVPGGAFVQALISHPTQITFTSTSTSTSTSTSTSTST 241
QY 240 TAARKNTRKVLGCKKAKAVMGLADLELAVKATVDSRVNSGGVNCAPVYVVGKICVDO 299
DB 242 AAKNLKKVITLDSKSPVIVFIAQIDKAVEPILVGFAGFNAQVCIAPSRLLIVHRSIID 301
QY 300 QFVNPLGEMAVG-EGNPAERNDIAMGLINAAALERVEQKVARAVEGAVFAGKAVFAGKAV 358
DB 352 EFVEKLEKVKKKLIGELFUSGNIYGPLISLQVDFPVI-SYTFE-KFEGAKVILGAFED 361
QY 359 ESK-----GYYPPLILLDVKOEMSMHETFGVPLVPAFTLELAISMANSDGLTSS 413
DB 362 ESKEYLGGYVQPTITDVTDMKMKFFTPGVPVLPILKEKDLDEALIELANDTEYGLAA 421
QY 434 SIYTCN-LNVAMKAIKGIKFGFTYINFPENFAMQ---SPHASMFKSTGGA-DKRLHLHG 468
DB 432 YVFETKILAPAFPAKALAEKIVWVNVV-VHAARFQLPFGGVVQSSNTEFHFSGKYKYLE 491
QY 469 YLQTVVYLO 478
DB 482 YTEIKTVIP 491

RESULT 5

US-10-175-636-21
Sequence 21, Application US10175696
Publication No. US20030092658A1
GENERAL INFORMATION:
APPLICANT: Glucksmann, Maria Alexandra
APPLICANT: Meyers, Rachel
APPLICANT: Rudolph-Owen, Laura A.
TITLE OF INVENTION: NOVEL HUMAN ENZYME FAMILY MEMBERS AND USES THEREOF
FILE REFERENCE: 10448-193001
CURRENT APPLICATION NUMBER: US10/175,696
PRIOR FILING DATE: 2002-06-20
PRIOR APPLICATION NUMBER: 10/067,668
PRIOR FILING DATE: 2002-02-04
PRIOR APPLICATION NUMBER: 60/256,140
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 09/842,901
PRIOR FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: 10/075,001,182,49
PRIOR FILING DATE: 2001-04-02
PRIOR APPLICATION NUMBER: 60/193,920
PRIOR FILING DATE: 2000-03-31
PRIOR APPLICATION NUMBER: 09/862,658
PRIOR FILING DATE: 2001-05-21
PRIOR APPLICATION NUMBER: PCT/US01/16380
PRIOR FILING DATE: 2001-05-21
PRIOR APPLICATION NUMBER: 60/205,675
PRIOR FILING DATE: 2000-05-19
PRIOR APPLICATION NUMBER: 09/882,837
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: PCT/US01/19319
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: 60/211,727
NUMBER OF SEQ ID NOS: 31
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 21
LENGTH: 493
TYPE: PRT


```
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12657

Query Match      31.5%; Score 771.5; DB 10; Length 496;
Best Local Similarity 36.4%; Pred. No. 6.2e-59;
Matches 170; Conservative 90; Mismatches 200; Indels 7; Gaps 3;

QY 10 YIDGQFVWPGDAMIDVNVNFAEAVISKIPGSAEAKAPKAIDAAEFA--QPEWEALPAIE 67
DB 12 YIDGWEVSANKNRDRIINYNGEVIPTVSEGTETAEAPPAIAARPAFESGEMSQETAET 71
QY 68 RASMLPKLSAIPKEKASRISALIVEEGRKIQGLAEVAVFAFYIYMAEWAHKKYGRFII 127
DB 72 RGKKVPAIADKIKHRRALAPLETITDGTGKTLSESYADMDIHNVMVYFACIADKDCGEM 131
QY 128 QSDRPPGENIILFKKALGVTTGILPNWPPFFLIAPKMAPALLTGTNTIVIKPSETTNNATA 187
DB 132 DSPIDPTESKIVKEPVGVVTOITPNYPLLOQASWKIAPALATGCSLVMKPSRITPLTTIR 191
QY 188 FAKIVDEIGLPRGVFNILGRGETVGOELAGNPKVAMVMTGVSAGEKIMATAAKNITK 247
DB 192 VFELMEEVGFPKGTINILGAGSEVGDVMSGHKEVDLVSTGTGTETCKHIMKNAANVTN 251
QY 248 VCLELGGKAPAIWMDADLELAVKAIIVSHVINSQGVNNAEPPVYVQKGIYDQFVNRLGE 307
DB 252 IALELGGKKNPNIIFDDAGFELAVDQALNGGYFHAGQVCSAGSRILVUNSLKDKFEQALID 311
QY 308 AMQAVQGNPAERNDIAMGPIINAAALERVQKVARAVENAPVAFSGRAVE---GKGY 363
DB 312 RVKTKIKLNGGTA-DTEMGPVISTEHPNKIESYMDVAKAPSATIAVSKRPDPDDLDKDL 370
QY 364 YYPPTLLDVRQEMSIMHEETFGVLPVPAFDLEDAISMANDSDYGLTSSIVTONINVA 423
DB 371 FPEPVIITNCDSIKRIVEEVGFPVTVGEFTEQEAQLQANSLVGLAGAVFSKDLGKA 430
QY 424 MKAIKGLKFGETYINRENFEAMQGFHAGWKSGIGGADGKHGLHGYL 470
DB 431 ORVANKLKLGTWINDRHPYPAQAPWGGYKQSGIGRGLKRGLEEVYL 477

RESULT 8
US-09-815-242-12657
; Sequence 12657, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: FIITPA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/264,308
; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12657
; LENGTH: 496
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12657

Query Match      31.5%; Score 771.5; DB 10; Length 496;
Best Local Similarity 36.4%; Pred. No. 6.2e-59;
Matches 170; Conservative 90; Mismatches 200; Indels 7; Gaps 3;

QY 10 YIDGQFVWPGDAMIDVNVNFAEAVISKIPGSAEAKAPKAIDAAEFA--QPEWEALPAIE 67
DB 12 YIDGWEVSANKNRDRIINYNGEVIPTVSEGTETAEAPPAIAARPAFESGEMSQETAET 71
QY 68 RASMLPKLSAIPKEKASRISALIVEEGRKIQGLAEVAVFAFYIYMAEWAHKKYGRFII 127
DB 72 RGKKVPAIADKIKHRRALAPLETITDGTGKTLSESYADMDIHNVMVYFACIADKDCGEM 131
QY 128 QSDRPPGENIILFKKALGVTTGILPNWPPFFLIAPKMAPALLTGTNTIVIKPSETTNNATA 187
DB 132 DSPIDPTESKIVKEPVGVVTOITPNYPLLOQASWKIAPALATGCSLVMKPSRITPLTTIR 191
QY 188 FAKIVDEIGLPRGVFNILGRGETVGOELAGNPKVAMVMTGVSAGEKIMATAAKNITK 247
DB 192 VFELMEEVGFPKGTINILGAGSEVGDVMSGHKEVDLVSTGTGTETCKHIMKNAANVTN 251
QY 248 VCLELGGKAPAIWMDADLELAVKAIIVSHVINSQGVNNAEPPVYVQKGIYDQFVNRLGE 307
DB 252 IALELGGKKNPNIIFDDAGFELAVDQALNGGYFHAGQVCSAGSRILVUNSLKDKFEQALID 311
QY 308 AMQAVQGNPAERNDIAMGPIINAAALERVQKVARAVENAPVAFSGRAVE---GKGY 363
DB 312 RVKTKIKLNGGTA-DTEMGPVISTEHPNKIESYMDVAKAPSATIAVSKRPDPDDLDKDL 370
QY 364 YYPPTLLDVRQEMSIMHEETFGVLPVPAFDLEDAISMANDSDYGLTSSIVTONINVA 423
DB 371 FPEPVIITNCDSIKRIVEEVGFPVTVGEFTEQEAQLQANSLVGLAGAVFSKDLGKA 430
QY 424 MKAIKGLKFGETYINRENFEAMQGFHAGWKSGIGGADGKHGLHGYL 470
DB 431 ORVANKLKLGTWINDRHPYPAQAPWGGYKQSGIGRGLKRGLEEVYL 477

RESULT 9
US-09-815-242-13829
; Sequence 13829, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: FIITPA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/264,308
; PRIOR FILING DATE: 2001-02-16
```

```

; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-23
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13829
; LENGTH: 481
; TYPE: PRT
; ORGANISM: Salmonella typhi
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(481)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-13829

```

```

Query Match      31.4%; Score 768.5; DB 10; Length 481;
Best Local Similarity 38.0%; Pred. No. 11; 58;
Matches 179; Conservative 75; Mismatches 204; Indels 13; Gaps 6;

```

```

QY 5 VQHPMYIDGQVTVWRGDAMTVVNPATPAVISPDPGQAFAPKAIIDAAEAQAQEWELP 64
DB 1 MQLQLLVGVLDGGERQ-SVYNPATGEVLEIAEASPVVLAALAAUSAFAPWQGT 66
QY 65 AIERASWLRKISAGIRERASEISALIVEEGSK-IQOLAEVEVAFTADYIDYMAEWARPYE 123
DB 67 PKRAECGLKLDSTFQNA:ETARIESQNGKPLHC-VINCEIPAIV:VEREFAGAAKPLS 126
QY 124 GEITQSDRPGENILLEKRALGVTTGILPNWPFELIARMMAPALLTGTIVIKSEFTN 183
DB 127 GLAAGEVLEHTSMIRDRPIGVVASIAPWNYPLMAAKLAPALAAAGNCVVIKSEFTPL 186
QY 184 NATAFAKIVDEIGLPGVFNVLGRGETVGOELAGNPKVAMVSMTGSVAGEKIMATAAK 243
DB 187 TALKLAALAKDI-FPGVLNVLFGRQTVGDVLTGHEKVRWVSLTGSITAGEHLKHTAP 245
QY 244 NITKVCLGEGKAPATVMDADLEIAVKAIVDSRVINSQVNCNCAERYVVKGYDQFVN 303
DB 246 AIKRTHMELGKAPVIVFTGAILAVAGQVRFTEGYNA-2GQ-IAA-RIVAGRGYDALVE 305
QY 304 KLCEAMQAVGFGPAERNIDAMGPLINAAALPVEQKVAPAVEG-ARVAFGGKAV 358
DB 306 KLGNVSSILKMGAPDKS-TEIGPISLTAHLR-----VTAVERAKALSHIPVITGSGQT 360
QY 359 EKGYYYPPTLLIDVQEMSINHEFTFGVLPVPAVFTGLEIAISMANSEVGLTSSIVTQ 418
DB 361 EKGYYVFAPTLLADAKQEDAIVQREVEGPPVSTIVEDDQGVLRWANISRYGLASSVWTQ 420
QY 419 NLAVAMKAIKGLKKEGYTYINRNEFEAMQGFHAGWKKSGTGGATGKHGLHGY 479
DB 421 DVXRAHLSARLQYGCITWINTHMLVSEMPHGGQKQSGYGDMSLYGLEUY 471

```

```

RESULT 10
US-09-815-242 10550
; Sequence 10550, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsson, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA-011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21

```

```

; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: 60/254,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,408
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10550
; LENGTH: 496
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10550

```

```

Query Match      31.2%; Score 762; DB 10; Length 496;
Best Local Similarity 35.0%; Pred. No. 4; 58;
Matches 168; Conservative 98; Mismatches 206; Indels 8; Gaps 5;

```

```

QY 3 VVQHPMYIDGQVTVWRGDAMTVVNPATPAVISPDPGQAFAPKAIIDAAEAQAQEWELP 62
DB 10 VASSYQLYINGEWITGSGNKMIASYNSNGEKALAEVIAINADIVRAVEAAAEALQWKO 69
QY 63 LPAIERASWLRKISAGIRERASEISALIVEEGSK-IQOLAEVEVAFTADYIDYMAEWAR 121
DB 70 VIWVTHSNITLKTATGTFENQFHLAMVFITDNGKPERFTQSDIVIASAHFERKFAVIRG 129
QY 122 YGEITQSDRPGENILLEKRALGVTTGILPNWPFELIARMMAPALLTGTIVIKSEFT 181
DB 130 EKSUVKEFKDQLSVV-KEPIGVVGQVILPNWPFELIARMMAPALLTGTIVIKSEFT 188
QY 182 INNATAFAKIVDEIGLPGVFNVLGRGETVGOELAGNPKVAMVSMTGSVAGEKIMATA 241
DB 189 SLSLELLEFLEGV-LPGGVNLTITGRGSDNSNYMLAHP-TPKLAFTGSGTGWGVTAFAA 247
QY 242 AKNITKVCLGEGKAPATVMDADLEIAVKAIVDSRVINSQVNCNCAERYVVKGYDQF 301
DB 248 ADPLPATLITGKSNATIFPDANWBERALRVGLGLINQGVVGNISRVVQSSITVQF 307
QY 302 VNFLEGEAMQAVGFGPAERNIDAMGPLINAAALPVEQKVAPAVEG-ARVAFGGKAV 359
DB 308 VEALKERTQVNVAGVTEWK-IWEMDAQINBQLEELKAYVELGKKEGATLITGGQREIN 366
QY 360 --GKYYYPPTLLIDVQEMSINHEFTFGVLPVPAVFTGLEIAISMANSEVGLTSSIVTQ 417
DB 367 GLKGAFLAPLILANGINIMVAQEEDEGVATVIRKFEFEVRLANGISEYGLGAVES 426
QY 418 QNLNVAMKAIKGLKKEGYTYINRNEFEAMQGFHAGWKKSGTGGATGKHGLHGY 477
DB 427 QHINVALPVAPGVPTSPKMWNTYN-QPAVALPQVYFFKSGTGFTHKSMILATLMMNYI 486

```

```

RESULT 11
US-09-815-242-12475
; Sequence 12475, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsson, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA-011A

```

```

; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/141,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12375
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12375

```

```

Query Match 30.98; Score 756.5; DR 10; Length 475;
Best Local Similarity 37.23; Pred. No. 1.2e-57;
Matches 175; Conservative 90; Mismatches 194; Indels 11; Gaps 7;

QY 10 YIDGQFTWRGDWIDVWNPATBAVISRIPDQGAEDARKAIDAARPAQFEWALPAIERA 69
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8 YINGEVESNSNETIEVINPATEEVIGKAKGNKADVDKAVEAADNVILEFRHTSVKREQ 67

QY 70 SWLPKISAGIPERASEISALIVEGGKIOQLAE-VEVAFTADYIDYMAEWARRYEGE 128
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 68 ALLKUIYKEVENKGGIVQAITDELGAPLSLSERVHYQGLNHFVAARKALLNYEFE--- 124

QY 129 SDRPGENILFKRALGVTTGILPWNFPFFLIARKMAPALLTNTIVIKPSEPTINNA 188
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125 -ERRGDD-LVVKEAIGVSLITPWNFPNTJSLKLAFAAGSPVVLKPSSETPFAAVIL 182

QY 190 AKIVETGIPWGVFNIIVLPGCEIVQGLACNPKVAMVMTGVSAGCFKMATAAKNTTKV 248
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 183 AEIFDKVGIVKGVFNIIVLPGCEIVQGLACNPKVAMVMTGVSAGCFKMATAAKNTTKV 242

QY 249 CIRLGGKAPATVMDADNLELAVKAIVDSRVINSQVNCAPRVVVKGIYDQFVNP 308
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 243 SLELGKSPYIVLDDVDVKEAKATTKCKVNNNTGQVCTAGTRVLPENKIKDAFLAELEQ 302

QY 309 MCAYQPNPAPNPIAMGPINAAALERVFKQKAVAFEGARVAFGSKAVGKGY 365
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 303 FSQVEVGNPRP-EGTQWSPITSKKQFTQGNVINKGIEETAFILFYNSPKRGLKGYFA 361

QY 366 PPTLLDVRQEMSTHMETFGPVLPPVAFDTLEDAISMANDSDYGLTSSITTONLVAMK 425
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 362 PPTIFINVDQMTIAQEEIFGPMVSVITYNDLDEATQIANDTKYGLAGYVIGKDETLHK 421

QY 426 AIKGLKFGETYINRNFENFAMGP-FAHWPKNSTLGGAPKHHGHCYLAQTQVYL 475
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 422 VAE-SLEAGHVLINAGKPKFLUFF QYKQSGGLGKENGXDGLDEEFLEK6SI 470

```

```

RESULT 12
US-09-344-882-22
; Sequence 22, Application US/09/444882
; Patent No. US20030162137A1
; GENERAL INFORMATION:
; APPLICANT: Nikolau, Basil J
; APPLICANT: Wurtele, Eve S
; APPLICANT: Oliver, David J
; APPLICANT: Behal, Robert
; APPLICANT: Schnable, Patrick S
; APPLICANT: Ke, Jinshan
; APPLICANT: Johnson, Jerry I

```

```

; APPLICANT: Alfred, Carolyn C
; APPLICANT: Fatland, Beth
; APPLICANT: Lutziger, Isabelle
; APPLICANT: Wee, Tsui-Jung
; TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
; FILE OF INVENTION: Acetyl CoA Levels in Plants
; FILE REFERENCE: 201573
; CURRENT APPLICATION NUMBER: US/09/344,882
; CURRENT FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 60/090,717
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Arabidopsis Thaliana
US-09-344-882-22

```

```

Query Match 30.63; Score 748.5; DR 9; Length 501;
Best Local Similarity 34.53; Pred. No. 6.3e-57;
Matches 165; Conservative 95; Mismatches 203; Indels 15; Gaps 6;

QY 9 MYIDGQFTWRGDWIDVWNPATBAVISRIPDQGAEDARKAIDAARPAQFEWALPAI 66
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22 LFINGQFUDASCKTETIOPRNGCEVIATIEGCKEULVLAVNAAAPYAFDHCQPMPTGF 81

QY 67 ERASWLKISAGIPERASEISALIVEGGKIOQLAE-VEVAFTADYIDYMAEWARRYEGE 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 82 ERAKLINKFADLIEENIEELAKLDVYGGKLFQGLKYADIPATAGHFRYNAGAADKHGE 141

QY 126 IIOSDRPGENILFKRALGVTTGILPWNFPFFLIARKMAPALLTNTIVIKPSEPTINNA 185
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 142 TLKWTQSLGTYTLKPEIGVGNLIIPWNEFSIMFATKVPAPAMAAGCTVVKPAEQTSLSA 201

QY 186 IAFAKIVDEITGLPRGVNVLVLRGETIVGOELACNPKVAMVMTGVSAGCFKIM-ATAAKN 244
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 202 LFVAHLKSKKAN:PNGVIN:VIGRSTAGAAATASHMIV:KVSFTQSTGVSKKMQAAAAASN 261

QY 245 ITVCLELGKAPATVMDADNLELAVKAIVDSRVINSQVNCAPRVVVKGIYDQFVNP 304
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 262 LKVSLEHDEKSPILLIPNADHAKAAL:QTFYNNRFP:VASSFVFWPFGTYNNKVVFP 321

QY 305 LGEMAOVQPNPAPNPIAMGPINAAALERVFKQKAVAFEGARVAFGSKAVGKGY 364
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 422 LVFPKAKMTWVQFPD-SIAP-GLV:KPK:FEKILSVIEH:KNEGAFH:GGSKAIGDKGYF 380

QY 365 YPPTLLDVRQEMSTHMETFGPVLPPVAFDTLEDAISMANDSDYGLTSSITTONLVAM 424
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 381 IQPTIEADVTFDMKTYQDEIFGPMVSVITYNDLDEATQIANDTKYGLAGHLSQDIDLIN 440

QY 425 KAIKGLKFGETYINRNFENFAMGP-FAHWPKNSTLGGAPKHHGHCYLAQTQVYL 477
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 441 TVSPSIKAGIITWVN-----GYHGFIDLCPYGYGYKMKSGNCFESOMATLNYITGTSVVM 493

```

```

PFSULT 13
US 15 223 865-22
; Sequence 22, Application US/10393865
; Publication No. US20030106090A1
; GENERAL INFORMATION:
; APPLICANT: Nikolau, Basil J
; APPLICANT: Wurtele, Eve S
; APPLICANT: Oliver, David J
; APPLICANT: Behal, Robert
; APPLICANT: Schnable, Patrick S
; APPLICANT: Ke, Jinshan
; APPLICANT: Johnson, Jerry L
; APPLICANT: Alfred, Carolyn C
; APPLICANT: Fatland, Beth
; APPLICANT: Lutziger, Isabelle
; APPLICANT: Wee, Tsui-Jung
; TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and

```

1 TITLE OF INVENTION: Acetyl CoA Levels in Plants
2 FILE REFERENCE: 217113
3 CURRENT APPLICATION NUMBER: US/10/293,865
4 PRIOR FILING DATE: 2002-11-13
5 PRIOR APPLICATION NUMBER: US 09/344,882
6 PRIOR FILING DATE: 1999-06-25
7 PRIOR APPLICATION NUMBER: US 60/050,717
8 PRIOR FILING DATE: 1998-06-26
9 NUMBER OF SEQ ID NOS: 38
10 SOFTWARE: PatentIn Ver. 3.1
11 SEQ ID NO 22
12 TYPE: PRT
13 LENGTH: 501
14 ORGANISM: Arabidopsis Thaliana
15 US-10-293-865-22

Query Match 30.6%; Score 748.5; DB 9; Length 501;
Best Local Similarity 34.5%; Pred. No. 6,3e-57;
Matches 165; Conservative 95; Mismatches 203; Indels 15; Gaps 6.

QY 9 MYIDGQVTVWRGDWIDVNVNPAVEAVISIPDQQAEDARKAIDAERA--QPEWEALPAI 66
Db 22 LFINGOFIDUASCKTETIDPNGEVATIAEHKEDVIAVNAAPYAFIHGSPWPMPTGPF 81
QY 67 ERASWIKKISARIRFRPASFISALIVEEGGKIQQLAE-VEVAFTADYIDYMAEWAFFRYEGE 125
Db 82 ERAKLINKEADALFEENIEELAKLDVADGKLPQKGYADIPATAGHFRYNAGAAKDTHGE 141
QY 126 IIOSDRPGENILFKKALGVYTGILPWNPFFLIARKMAPALLTNTIVIKPSFETNNA 185
Db 142 TLKMTKOSLPGYTLKEPIGPGWGNITPWNPFSIMFATKVPAPAMAAGCTMVVKPAPQTSLSA 201
QY 186 IAFKTIKVTGIGLPGGVNVLVGRGETVGGELAGNPKVAMVSMTGSVSAKIM-ATAAKN 244
Db 202 LFVAHLSKEAGIPDGVNLVITGFGTAGAAIAASHMDVDKVSFTGSTDVGRKIMQAAASN 261
QY 245 ITKVCLELGGKAPAVIMDDADLELAVKAIVDSRVINSOVNCNCAERVYVQKGIYDQFVNR 304
Db 262 LKRVSLGKSPKLLIFNDADIDKAADLALLGCFYKNGEICVASSKVFVQVGIYDKVVEK 321
QY 305 LGAMQAVQGNPAERNDIAMGPLINAAALERVEQKVARAVEEGARVAFGKAVGKGY 364
Db 322 LVKAKMTVQDFD-STARQSPQVUKKQFELSYIEHGNEGATILTGKALIGDKGYF 380
QY 365 YPPTLLDVKRQESIMHEETFGFVLPVAVFDLEDAISMANDSDYGLTSSITVONLNVAM 424
Db 381 IOPTIEADVTEDMKIYODEIFGPNVSLMKFKTVEEGTKCANNTKYGLAAGILSQDIDLIN 440
QY 425 KALKGLKPGHTYINRENPFAMQGF-----HACWRKSGIGGAGDKHGLHGYLOTQVVYL 477
Db 441 TVSRIRKAGILWVN-----CYFGFDLCPYGGYKWSGNCRESGMDALDNYLQTKSVVM 493

RESULT 14
US-09-748-626-3558
Sequence 3558, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAKO
APPLICANT: SENOH, AKIHIO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
PRIOR FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484

1 PRIOR FILING DATE: 1999-12-16
2 PRIOR APPLICATION NUMBER: JP 00/159162
3 PRIOR FILING DATE: 2000-04-07
4 PRIOR APPLICATION NUMBER: JP 00/280988
5 PRIOR FILING DATE: 2000-08-03
6 NUMBER OF SEQ ID NOS: 7059
7 SOFTWARE: PatentIn ver. 3.0
8 SEQ ID NO 3558
9 LENGTH: 490
10 TYPE: PRT
11 ORGANISM: Corynebacterium glutamicum
12 US-09-738-626-3558

Query Match 29.8%; Score 729; DB 9; Length 490;
Best Local Similarity 34.4%; Pred. No. 4.1e-55;
Matches 160; Conservative 89; Mismatches 206; Indels 10; Gaps 4.

QY 20 GDAWI-----DVVNPAEAVISRIPIQQAERHARKAIDAERACQPEWEALPAERASW 71
Db 19 GDSWVEASDGGTFDVENPATGETIATLASATSEIALAALDAACAVJALWAPAPAKRSNI 78
QY 72 LFKISWCIPEPASEISALIVEEGGKIQQLAEVFAFTADYIDYMAFWAPPYEGELIJSUR 141
Db 79 LRKGFELVAERAREEATILTEEMKELARAGEVTVNNEFLRWFSEFAVRILYGRYGIPE 148
QY 132 PCENILLPKALGVYTGILPWNPFFLIARKMAPALLTNTIVIKPSFETNNAIATAKI 191
Db 139 GNLRLMTALKPVGPCLLITPWNPFLAMATRKVAVAIAGVVMVLKPAKLPILTSQYFAGT 198
QY 192 VIEGELPGGVNVLVGG-EGETVGGELAGNPKVAMVSMTGSVSAKIM-ATAAKRITKVT 250
Db 199 MLIAAGLPAPVNLVNSASASALNPIWMEHGRKVSFDSIPV-GLTIRKKAALKAVIRISM 258
QY 251 FICGKAPAVIMDDADLELAVKAIVDSRVINSOVNCNCAERVYVQKGIYDQFVNR 310
Db 259 ELGGNAPFVFEEDADILATGACGAKKRNIGEA-TAARFLVHSVADEPRERFAAKLE 318
QY 311 AVGEPNPAERNITAMPD-INAAALFEPVQKVARAVEEGARVAFGKAVGKGYIYDQFVNR 370
Db 319 EGVLANGLDEG-VTVGPIVERKARSDVASIVIAAVAKGATVLTGCKATGAGYFYEPTVL 377
QY 371 IIVYQPMSTIMHEETFGFVLPVAVFDLEDAISMANDSDYGLTSSITVONLNVAMKAIKGL 400
Db 378 TGVSTUADAILNEIEFGFVLPVAVFDLEDAISMANDSDYGLTSSITVONLNVAMKAIKGL 437
QY 431 KKEGELYINKENE-EMAGGPHAMWKKSGELGAGKKEHLEHAYLOTQVV 474
Db 438 EPCILVNSGVISNAAALFPGVYKWSGNCRESGMDALDNYLQTKSVYL 492

RESULT 15
US-09-815-242-12102
Sequence 12102, Application US/00015242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: FLTRA-011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: 60/207,323

PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO: 12102
LENGTH: 490
TYPE: PRI
ORGANISM: Pseudomonas aeruginosa
US-09-815-242-12102

Query Match 29.8%, Score 729, DB 10, Length 490;
Best Local Similarity 34.2%, Pred. No. 3, le-55;
Matches 164, Conservative 101, Mismatches 206, Indels 8, Gaps 5;

QY 6 QHPYIDGQVYTWBGDAMIDVNPATPAVISRIIDGQEDARKAIDAERAOPMEALPA 65
DB 6 EQKLYIGGRVYEASSGAFELINPANGEVLAKEVAKASKEDEVAKAVSAVEGQKVMAMJA 65
QY 66 IEFASWIKKISAGIIEFASSELALIVEE3K-IQJLAEEVATADYIDYMAEMARREES 124
DB 66 MORSPTIIPPAVDILPEPNDLAELETLDTSKPLAETPSVDITVGADVLEYYAGLVPAIEG 125
QY 125 EIIOSDRPGENILFEKRALGVTTGILPWNPEFLARKMAPALLTNTIYIKPSEPTNN 184
DB 126 EQIPL-RETSFYTRREPLGVAGIGAMNYPVQIALMKSPALAGNAMIFKPEVTPLT 184
QY 185 AIAFAKIVDELGLRGVFNVLIGRGETVGOELAGNPKVAVSMTGSVSGEKIMATA-AK 243
DB 185 ALKLAELYTEAGVDPGVFNLTGSGREVGOMLTHEPLIEKISFTGISTGKKVMASASS 244
QY 244 NITVVCLELGGKAPAIWDDADLELAVKATVDSRVINSQOVNCAERVYVQKGIYDQFN 303
DB 245 SLKEVTMLGSKSPILIFPDADLDRKADLIYMANFSSGVCTNGTIRVFIHRSQARFEA 304
QY 304 RIGAMQAVCPGNPAFRNDIAMPPTINAAMIFVEQKVARAVEGARVAFGKAVE---- 359
DB 305 KLEERVCRIRIGDPQDEN-TNPGPLVSFPMESVLYTESGKAQKALLGGGERVUDGAF 363
QY 360 GKGYIYPTLLLDLVKQESLMHEETRGFVLPEVAEPDLEDAISMANDSDYGLTSSITYON 419
DB 364 GKGAVALPPTVTPCPDMITVPRIFGPVMSIIVYDDEDAIRANDTEYGLAGVYTD 423
QY 420 INVAKKAKIKIKKEEIVINFEFNFAWGPFAWPKSITGSAIKKHSIASHYIATVVVYIQ 479
DB 424 LARAHRAIHRIEAGICINNTMGESPALMPYGVGKQSGVGENGLITLAHTRIKSYOVE 482

Search completed: June 24, 2003, 10:37:06
Job time: 24.1541 secs

[The body of the page contains extremely faint, illegible text, likely bleed-through from the reverse side of the document. The text is organized into several paragraphs, but the characters and words are not discernible.]

Genome version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: Tue Jun 24, 2003, 10:11:00 : Search time 17 x 414 seconds

(without alignments)
1780.862 Million cell updates/sec

Title: us-09-830-751-6

Sequence score: 2446
1 MSVPVHPMTLDGEVFWB3...ATGKH5LH3YLUYVYU5 479

Scoring table: BLASTSUM62

Gapop 10.0, Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

```

1 Pending Patents-AA-Main.*
2 /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
3 /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
4 /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
5 /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
6 /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
7 /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
8 /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
9 /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
10 /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
11 /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
12 /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
13 /cgn2_6/ptodata/1/paa/US097_COMB.pep.*
14 /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
15 /cgn2_6/ptodata/1/paa/US099_COMB.pep.*
16 /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
17 /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
18 /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
19 /cgn2_6/ptodata/1/paa/US103_COMB.pep.*
20 /cgn2_6/ptodata/1/paa/US104_COMB.pep.*
21 /cgn2_6/ptodata/1/paa/US105_COMB.pep.*
22 /cgn2_6/ptodata/1/paa/US106_COMB.pep.*
23 /cgn2_6/ptodata/1/paa/US107_COMB.pep.*
24 /cgn2_6/ptodata/1/paa/US108_COMB.pep.*
25 /cgn2_6/ptodata/1/paa/US109_COMB.pep.*
26 /cgn2_6/ptodata/1/paa/US110_COMB.pep.*
27 /cgn2_6/ptodata/1/paa/US111_COMB.pep.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2432	99.4	479	US-09-791-537-112887	Sequence 112887, App1
2	2432	99.4	479	US-09-801-563-9	Sequence 9, App1
3	2432	99.4	479	US-09-360-039-843	Sequence 823, App
4	2157	88.7	514	US-09-489-039A-13820	Sequence 13820, A
5	1557	63.7	480	US-09-791-537-40455	Sequence 40455, A
6	1552	63.5	480	US-09-791-537-115281	Sequence 115281, A

Query Match	Score	DB ID	Length	App
99.4%	2432	US-09-791-537-112887	479	App1
99.4%	2432	US-09-801-563-9	479	App1
99.4%	2432	US-09-360-039-843	479	App
88.7%	2157	US-09-489-039A-13820	514	A
63.7%	1557	US-09-791-537-40455	480	A
63.5%	1552	US-09-791-537-115281	480	A

ALIGNMENTS

```

RESULT 1
US-09-791-537-112887
Sequence 112887, App1:391-1500 95...3791537
GENERAL INFORMATION:
APPLICANT: Biocompare, Inc.
APPLICANT: Danzer, Joseph
TITLE OR INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
FILE REFERENCE: 251/210
CURRENT APPLICATION NUMBER: US/09/791,537
CURRENT FILING DATE: 2001-02-22
NUMBER OF SEQ. IN N.S.: 153055
SOFTWARE: Patent version 3.0
SEQ ID NO 112887
LENGTH: 479
TYPE: PRT
ORGANISM: Escherichia coli
US-09-791-537-112887

```

Query Match	Score	DB ID	Length	App
99.4%	2432	US-09-791-537-112887	479	App1
99.4%	2432	US-09-801-563-9	479	App1
99.4%	2432	US-09-360-039-843	479	App
88.7%	2157	US-09-489-039A-13820	514	A
63.7%	1557	US-09-791-537-40455	480	A
63.5%	1552	US-09-791-537-115281	480	A

```

Db      61  EALPAIERASMLRKISAGIREPASRISALIVEEGCKTQGLAFVFAFTADYIDYAEAR 120
      121  KYEPEIIQSDRPGENILLEKRALGVTTGILPMNPFPLLARKMAPALLTGNITVIRKPEEF 180
      121  KYEPEIIQSDRPGENILLEKRALGVTTGILPMNPFPLLARKMAPALLTGNITVIRKPEEF 180
      181  TTNNALIAFAKIVDEIGLPGVFNILVLSKGETVGGELAGNPKVAMVSMGTSVSAEKMAT 240
      181  TTNNALIAFAKIVDEIGLPGVFNILVLSKGETVGGELAGNPKVAMVSMGTSVSAEKMAT 240
      241  AAKNITKVCLELGKAPALVMDADLELAVKATVDSRVINSQGVNCAKRVYVOKGTYDQ 300
      241  AAKNITKVCLELGKAPALVMDADLELAVKATVDSRVINSQGVNCAKRVYVOKGTYDQ 300
      401  FVNRLGEAMQAVQFGNPAERNDIAMGPLINAAALEVEBOKVAVAEBAVAFGKRAVEG 360
      401  FVNRLGEAMQAVQFGNPAERNDIAMGPLINAAALEVEBOKVAVAEBAVAFGKRAVEG 360
      461  KGYVYPTLLIDVROKMSIMHEETPGVLPVVAFTLEDAISMANDSYGITSSITYONL 420
      461  KGYVYPTLLIDVROKMSIMHEETPGVLPVVAFTLEDAISMANDSYGITSSITYONL 420
      421  NVAMKAIKGLKFGETYINRENFAMOGFHAGMKRSIGCAJCKHGLHETLOTQVYVYLOS 479
      421  NVAMKAIKGLKFGETYINRENFAMOGFHAGMKRSIGCAJCKHGLHETLOTQVYVYLOS 479

```

RESULT 2

```

US-09-801-563-9
: Sequence 9, Application US/09801563
: GENERAL INFORMATION:

```

```

: APPLICANT: Levy, Stuart, et. al.
: TITLE OF INVENTION: NIMH COMPOSITIONS AND THEIR METHODS OF USE
: FILE REFERENCE: PKZ-043
: CURRENT FILING DATE: 2001-03-08
: PRIOR APPLICATION NUMBER: 60/188,362
: PRIOR FILING DATE: 2000-04-10
: NUMBER OF SEQ ID NOS: 98
: SOFTWARE: Paton In-Vit. 2.0
: SEQ ID NO 9
: LENGTH: 479
: TYPE: PRT
: ORGANISM: Escherichia coli
US-09-801-563-9

```

Query Match

```

Best Local Similarity 99.4%; Score 2432; DB 27; Length 479;
Matches 477; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

      1  MSYVQHPMTIDGQFTWRGDAMIIVVNATIAVLSRIPDGAEDARAKAIDAERAQPEW 60
      1  MSYVQHPMTIDGQFTWRGDAMIIVVNATIAVLSRIPDGAEDARAKAIDAERAQPEW 60
      61  EALPAIERASMLRKISAGIREPASRISALIVEEGCKTQGLAFVFAFTADYIDYAEAR 120
      61  EALPAIERASMLRKISAGIREPASRISALIVEEGCKTQGLAFVFAFTADYIDYAEAR 120
      121  KYEPEIIQSDRPGENILLEKRALGVTTGILPMNPFPLLARKMAPALLTGNITVIRKPEEF 180
      121  KYEPEIIQSDRPGENILLEKRALGVTTGILPMNPFPLLARKMAPALLTGNITVIRKPEEF 180
      181  TTNNALIAFAKIVDEIGLPGVFNILVLSKGETVGGELAGNPKVAMVSMGTSVSAEKMAT 240
      181  TTNNALIAFAKIVDEIGLPGVFNILVLSKGETVGGELAGNPKVAMVSMGTSVSAEKMAT 240
      241  AAKNITKVCLELGKAPALVMDADLELAVKATVDSRVINSQGVNCAKRVYVOKGTYDQ 300
      241  AAKNITKVCLELGKAPALVMDADLELAVKATVDSRVINSQGVNCAKRVYVOKGTYDQ 300
      401  FVNRLGEAMQAVQFGNPAERNDIAMGPLINAAALEVEBOKVAVAEBAVAFGKRAVEG 360
      401  FVNRLGEAMQAVQFGNPAERNDIAMGPLINAAALEVEBOKVAVAEBAVAFGKRAVEG 360

```

```

Db      301  FVNRLGEAMQAVQFGNPAERNDIAMGPLINAAALEVEBOKVAVAEBAVAFGKRAVEG 360
      361  KGYVYPTLLIDVROKMSIMHEETPGVLPVVAFTLEDAISMANDSYGITSSITYONL 420
      361  KGYVYPTLLIDVROKMSIMHEETPGVLPVVAFTLEDAISMANDSYGITSSITYONL 420
      421  NVAMKAIKGLKFGETYINRENFAMOGFHAGMKRSIGCAJCKHGLHETLOTQVYVYLOS 479
      421  NVAMKAIKGLKFGETYINRENFAMOGFHAGMKRSIGCAJCKHGLHETLOTQVYVYLOS 479

```

RESULT 2

```

US-60-360-039-823
: Sequence 823, Application US/60360039
: GENERAL INFORMATION:

```

```

: APPLICANT: Cao, Yongwei
: APPLICANT: Chen, Xianfeng
: APPLICANT: Goldman, Barry S.
: APPLICANT: Hinkle, Gregory J.
: APPLICANT: Slater, Steven C.
: TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
: FILE REFERENCE: 36-10(52052)A
: CURRENT FILING DATE: 2002-02-21
: NUMBER OF SEQ ID NOS: 47374
: SEQ ID NO 823
: LENGTH: 479
: TYPE: PRT
: ORGANISM: Escherichia coli
US-60-360-039-823

```

```

Query Match
Best Local Similarity 99.4%; Score 2432; DB 27; Length 479;
Matches 477; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

      1  MSYVQHPMTIDGQFTWRGDAMIIVVNATIAVLSRIPDGAEDARAKAIDAERAQPEW 60
      1  MSYVQHPMTIDGQFTWRGDAMIIVVNATIAVLSRIPDGAEDARAKAIDAERAQPEW 60
      61  EALPAIERASMLRKISAGIREPASRISALIVEEGCKTQGLAFVFAFTADYIDYAEAR 120
      61  EALPAIERASMLRKISAGIREPASRISALIVEEGCKTQGLAFVFAFTADYIDYAEAR 120
      121  KYEPEIIQSDRPGENILLEKRALGVTTGILPMNPFPLLARKMAPALLTGNITVIRKPEEF 180
      121  KYEPEIIQSDRPGENILLEKRALGVTTGILPMNPFPLLARKMAPALLTGNITVIRKPEEF 180
      181  TTNNALIAFAKIVDEIGLPGVFNILVLSKGETVGGELAGNPKVAMVSMGTSVSAEKMAT 240
      181  TTNNALIAFAKIVDEIGLPGVFNILVLSKGETVGGELAGNPKVAMVSMGTSVSAEKMAT 240
      241  AAKNITKVCLELGKAPALVMDADLELAVKATVDSRVINSQGVNCAKRVYVOKGTYDQ 300
      241  AAKNITKVCLELGKAPALVMDADLELAVKATVDSRVINSQGVNCAKRVYVOKGTYDQ 300
      301  FVNRLGEAMQAVQFGNPAERNDIAMGPLINAAALEVEBOKVAVAEBAVAFGKRAVEG 360
      301  FVNRLGEAMQAVQFGNPAERNDIAMGPLINAAALEVEBOKVAVAEBAVAFGKRAVEG 360
      361  KGYVYPTLLIDVROKMSIMHEETPGVLPVVAFTLEDAISMANDSYGITSSITYONL 420
      361  KGYVYPTLLIDVROKMSIMHEETPGVLPVVAFTLEDAISMANDSYGITSSITYONL 420
      421  NVAMKAIKGLKFGETYINRENFAMOGFHAGMKRSIGCAJCKHGLHETLOTQVYVYLOS 479
      421  NVAMKAIKGLKFGETYINRENFAMOGFHAGMKRSIGCAJCKHGLHETLOTQVYVYLOS 479

```

RESULT 4

```

US-09-489-039A-13820
: Sequence 13820, Application US/09489039A
: GENERAL INFORMATION:

```

Query Match 63.78; Score 1557; DB 21; Length 480;

186 IAFKIVEIETFGVINEVADJEEI VDEIADIKVAMVMTGVNATFKIMATAKNI 245

2. TITLE OF INVENTION: CLONACE FOR DIAGNOSTICS AND THERAPEUTICS

[illegible]

```

Db      61 PWNPFPLIAPKMAPALLTGNMTVIKPSSEFTNNNAFEPAFLVSSQLPKGVFNLVACKCS 120
Qy      211 TVGOELAGPKVAMYSMTGSVAGEKIMATAKNTIVCLELGGKAPAIYMDADLELAV 270
      121 VVGELESENENIGWISLGSVEAGT RWEAAAKNI IVSLELEGGKAPAIYCKADIDLAV 180
Qy      273 KAVISPVINSQVNCFAERYVYQKJIDYFVNPLJSLAMQAVPCNIAFENULIAMGLIN 330
      181 EAKISKRLGNNGVNC ABRATVHTSVLEFVLCFKVAKMSKVSQNTL-KADEMGFLVN 239
Db      331 AAALREVDKARAAVEGARAFAFGKAVEGKYVPTLLIDVROEKSIMHEETGEFLP 390
      240 GAVGNALAMLOKAKAIKAIIVEGCKITITSGYVFPASVLINVKHELEIMOKELFALILP 299
Qy      391 VVAFTLEDAISMANDSYGLISSTYONLVANKAKIKLKEGTYINKENFEAMOGFHA 450
      300 IAKFTLDEVIDMADCEYGLTSSLYONLIDIAMRASHPEIKPPTVINPENFEAMOGFHA 359
Db      451 GWRKSGIGADGKHGHGLOTGVVYLD 478
      360 GFRKSGIGADGKHGHGLOTGVVYLD 367

RESULT 10
US-09-450-969-4782
: Sequence 4782, Application US/09450969
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: FILE REFERENCE: PAT99-09A
: CURRENT APPLICATION NUMBER: US/09/450,969
: PRIORITY FILING DATE: 1999-11-29
: NUMBER OF SEQ ID NOS: 7544
: SEQ ID NO 4782
: LENGTH: 493
: TYPE: PPT
: ORGANISM: S. epidermidis
US-09-450-969-4782

Query Match      38.1%; Score 932.5; DB 18; Length 493;
Best Local Similarity 38.5%; Pred. No. 3.8e-83;
Matches 181; Conservative 108; Mismatches 180; Indels 1; Gaps 1;

Qy      9 MYIDGFYTPGDADIDVNPATFAVISPITFGAPAPKATPAAPAPPEMATAPIEP 68
      24 LEINNEFLSOGKEMVYNPATGEPDTITLATEEVNDAIEKSCQAOJLMEHVRQPTK 63
Db      69 ASWLPKISAGIPEPASETSALTVERGKTLQGLAFVVAFTADYIDVMAEMARREGRITQ 128
      84 AERHVLITILEKNRDEIAQIVYVEGCTTLAGVGLIKSISFLDYMSLSMSKSPVLD 143
Qy      129 SRPPGPNITLPRKRAIGVTGTLPMNPFPLIAPKMAPALLTGNMTVIKPSSEFTNNNAIAF 188
      144 NSTIANETIQIINKPLGVTAGIVPMNAPILVIMKVIYLAIVGSSVILKPSSEFTLLILKL 204
Db      189 AKIVDEILPRGVENLVIRGETVGOELAGPKVAMYSMTGSVAGEKIMATAKNTIKY 248
      204 AELPFASTPATLFEIVGTGTEPTVTSGLASHKIDQLISLTSMAKVSVEENAAQIVKKV 264
Qy      249 LILVVGKENSIMHEETGEFLVVAAPDLEDAISMANDSYGLISSTYONLVANKAKIK 428
      264 NLELGGNNAVIVTSNARIDKAVNYIVTARINNAGQVTCPEPIFVHEVHUDEFNKVISK 323
Db      309 MAAVVGSPAEENULIAMSPILNAAALFEVGRKAVAFEGARVAFEGKAVEGKYVPT 368
      324 MSLIVGSPREPN-LDYGAILINQKQLSIRHEKVDIAKKNATIKTGGHOLKRHGFYAPT 382
Qy      369 LILVVGKENSIMHEETGEFLVVAAPDLEDAISMANDSYGLISSTYONLVANKAKIK 428
      383 VILNWKTYNVEKEFEFVLAIIYIDFEGVLELANIDNGLSSYIFSENLFEVMTATE 442
Qy      429 GLKFGFTYINPENFAMOGFHAQWMSKSIISGAPFIRHPEFYNTTYSYIR 478

```

```

Db      443 RLKFEVYANCBAEEVNGYHAGWBSQDAGDIGHFEFYNTTYSYIR 492

RESULT 11
US-09-710-279-2956
: Sequence 2956, Application US/09710279
: GENERAL INFORMATION:
: APPLICANT: KIMBERLY, WILLIAM JOHN
: TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
: FILE REFERENCE: P0348005
: CURRENT APPLICATION NUMBER US/09/710,279
: PRIORITY FILING DATE: 2000-11-09
: PRIOR APPLICATION NUMBER: 60/164,258
: NUMBER OF SEQ ID NOS: 4472
: SOFTWARE: Patent In Ver. 2.1
: SEQ ID NO 2956
: LENGTH: 479
: TYPE: PPT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: synthetic
: OTHER INFORMATION: amino acid sequence
US-09-710-279-2956

Query Match      38.1%; Score 931.5; DB 21; Length 479;
Best Local Similarity 38.5%; Pred. No. 4.6e-83;
Matches 181; Conservative 109; Mismatches 179; Indels 1; Gaps 1;

Qy      9 MYIDGFYTPGDADIDVNPATFAVISPITFGAPAPKATPAAPAPPEMATAPIEP 68
      10 LEINNEFLSOGKEMVYNPATGEPDTITLATEEVNDAIEKSCQAOJLMEHVRQPTK 69
Db      69 ASWLPKISAGIPEPASETSALTVERGKTLQGLAFVVAFTADYIDVMAEMARREGRITQ 128
      70 AERHVLITILEKNRDEIAQIVYVEGCTTLAGVGLIKSISFLDYMSLSMSKSPVLD 129
Qy      129 SRPPGPNITLPRKRAIGVTGTLPMNPFPLIAPKMAPALLTGNMTVIKPSSEFTNNNAIAF 188
      139 NSTIANETIQIINKPLGVTAGIVPMNAPILVIMKVIYLAIVGSSVILKPSSEFTLLILKL 189
Db      189 AKIVDEILPRGVENLVIRGETVGOELAGPKVAMYSMTGSVAGEKIMATAKNTIKY 248
      196 AELPFASTPATLFEIVGTGTEPTVTSGLASHKIDQLISLTSMAKVSVEENAAQIVKKV 249
Qy      249 LILVVGKENSIMHEETGEFLVVAAPDLEDAISMANDSYGLISSTYONLVANKAKIK 428
      250 NLELGGNNAVIVTSNARIDKAVNYIVTARINNAGQVTCPEPIFVHEVHUDEFNKVISK 409
Db      309 MAAVVGSPAEENULIAMSPILNAAALFEVGRKAVAFEGARVAFEGKAVEGKYVPT 368
      310 MSLIVGSPREPN-LDYGAILINQKQLSIRHEKVDIAKKNATIKTGGHOLKRHGFYAPT 368
Qy      369 LILVVGKENSIMHEETGEFLVVAAPDLEDAISMANDSYGLISSTYONLVANKAKIK 428
      429 GLKFGFTYINPENFAMOGFHAQWMSKSIISGAPFIRHPEFYNTTYSYIR 478
Db      429 GLKFGFTYINPENFAMOGFHAQWMSKSIISGAPFIRHPEFYNTTYSYIR 478

RESULT 12
US-60-360-039-18199
: Sequence 18199, Application US/60360039
: GENERAL INFORMATION:
: APPLICANT: Cao, Yongwei
: APPLICANT: Chen, Xiaofeng
: APPLICANT: Goldman, Barry S.
: APPLICANT: Hinkle, Gregory J.
: APPLICANT: Slater, Steven C.
: TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

```

Query Match	46.88;	Score 900;	DH 27;	Length 473;
Best Local Similarity	48.58;	Prod. No. 6, 2e 80,		
Matches 182;	Conservative 95;	Mismatches 188;	Indels 8;	Gaps 23

[illegible]

[illegible]

```

RESULT 15
PCT-US02-03987-5058
: Sequence 5058, Application PC/US0203987
: GENERAL INFORMATION:
: APPLICANT: Elittra Pharmaceuticals, Inc.
: TITLE OF INVENTION: Methods for identifying the Target of a Compound which Inhibits
: TITLE OF INVENTION: Proliferation
: FILE REFERENCE: ELITRA.028VPC
: CURRENT APPLICATION NUMBER: PCT/US02/03987
: CURRENT FILING DATE: 2002-02-02
: PRIOR APPLICATION NUMBER: 60/267,636
: PRIOR FILING DATE: 2001-02-09
: NUMBER OF SEQ ID NOS: 15811
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 5058
: LENGTH: 483
: TYPE: PRT
: ORGANISM: Pseudomonas aeruginosa
PCT-US02-03987-5058

```

Query Match	36.0%;	Score 879.5;	DB 1;	Length 483;
Best Local Similarity	39.3%;	Pred. No. 7.2e-78;		
Matches 184;	Conservative	96;	Mismatches 187;	Indels 1;
				Gaps 1

```

QY 1 YIDGCVTRMGAMIDIVVPATEAVTSIRPGQAEADAKALDAERAPQEWELPATERA 69
Db 14 YVDGAVVADNDNOTIKVNNPATGELIGSVPKMGAETRALEADKALPMARLITAKERA 73
QY 70 SWLRKTSAGIRERBASISALIVEEGCKTQOLAEEVAFADYIDYMAEMARREGETIOS 129
Db 74 NKLERFEMDIMEODDLARLMTIEQCKPLAEAKGEIAYAA5FLPMEPEEMARIYGDITPG 133
QY 130 DRGEMILLFKBALGVTTGILLPMNPFLLIARKMAPALLTGNTIVIRPSEFTNNALAF 189
Db 134 HQDPKRTIYKQJYGTVAALITPMNPSAMITPKKGPALAACTWLKASOTPYSAIALA 193
QY 190 K1VDELTGLPRVFNIYVLPGETVQGLAGNEKRVAMSVMSVSAGEKIMATAKNITKVC 249
Db 194 ELAERAGTIPKGVFSVTSAGRV3GSELT5NPIVTKLFRTGTSBELG2PLMA6CAQDIKVS 253
QY 250 LEL3GKAPALVMDALIELAVKAVIYNS3LVN7AEVYVYKQK1YIQVPHN1LGEAM 309
Db 254 LEL3GAPR1VFD0DALDAVEAGALISYPPNNQ3QVYVANPLVYDQYDAFYDKLEAAV 313
QY 310 QAVQFNP1PERKNDIMAGPLINAALEEREOUVARAVEGAVRGAVERGKAVBEQGYRPTL 369
Db 314 AKINIONGIEAG-VITGPLLAKAVAKKEEHADAVSGAVVSGCKPRLAGTFFPEPTI 372
QY 370 LLDVROEMSIMHEETGPVLVPAVAFDTLEDA1SMANDSDYGLT5IYT0NLNMAKAIK 429
Db 373 LVYDVPKNALVSRDETFEGPLAPVFRPKDEAEVIAMSNDTEFGCLASYFAAPDLAPVFAEQ 432
QY 430 LKRGCEYIINRNFEMOGFHCNMRKSGIGAGADKHGLHGYIQTOVYVL 477
Db 433 LEYGMGINTGLISNEVAPFGIKSGAGREGSKYGEDYELIKYUCL 480

```

1

2

GenCore version 5.1.6
Copyright (c) 1994 - 2003 Compugen Ltd

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:14:35, Search time 51.8555 Seconds
(without alignments)
2375.712 Million cell updates/sec

Title: US-09-830-751-6

Sequence: 1 MSVPVGHMPTITGQFVITWKA...ANKKH-LHRYLQIVLVVLS 479

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 1171708 seqs, 257189365 residues

Total number of hits satisfying chosen parameters: 1171708

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

Pending Patents: AA-New: *
1: /cgn2_6/plodata/2/paa/FCI_NEW_COMB.pep.*
2: /cgn2_6/plodata/2/paa/BS06_NEW_COMB.pep.*
3: /cgn2_6/plodata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/plodata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/plodata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/plodata/2/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/plodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2446	100.0	479	US-09-830-751-6	Sequence 6, Appl 1
2	2432	99.4	479	US-10-369-493-823	Sequence 823, App
3	2157	88.2	514	US-10-446-203-13820	Sequence 13820, A
4	1552	63.5	480	US-10-282-122A-65716	Sequence 65716, A
5	1499	61.3	319	US-10-417-HM6-5824	Sequence 5824, Ap
6	1279.5	52.3	393	US-10-282-122A-54340	Sequence 54340, Ap
7	932.5	38.1	493	US-10-092-411A-4388	Sequence 4388, Ap
8	903	36.9	493	US-10-369-493-18199	Sequence 18199, A
9	900.5	36.8	480	US-10-282-122A-65076	Sequence 65076, A
10	900	36.8	473	US-10-369-493-13571	Sequence 13571, A
11	879.5	36.0	483	US-10-282-122A-43455	Sequence 43455, A
12	879.5	36.0	594	US-10-365-683-27749	Sequence 27749, A
13	879.5	36.0	594	US-10-419-128-27749	Sequence 27749, A
14	879.5	36.0	594	US-10-419-128-27749	Sequence 27749, A
15	873.5	35.7	482	US-10-282-122A-56650	Sequence 56650, A
16	873.5	35.7	482	US-10-369-493-23511	Sequence 23511, A
17	868.5	35.5	482	US-10-282-122A-75217	Sequence 75217, A
18	863.5	35.3	478	US-10-369-493-13957	Sequence 13957, A
19	852.5	35.1	460	US-10-282-122A-16820	Sequence 16820, A
20	859.5	35.1	465	US-10-369-493-7938	Sequence 7938, Ap
21	859.5	35.1	482	US-10-282-122A-75920	Sequence 75920, A
22	853.5	34.9	482	US-10-282-122A-59544	Sequence 59544, A
23	853.5	34.9	487	US-10-446-203-13533	Sequence 13533, A
24	845.5	34.6	475	US-10-369-493-17412	Sequence 17412, A
25	836	34.2	483	US-10-282-122A-46932	Sequence 46932, A
26	835	34.1	468	US-10-369-493-17169	Sequence 17169, A

27	833.5	34.1	489	US-10-282-122A-51224	Sequence 51224, A
28	842	44.0	486	US-10-282-122A-77428	Sequence 67328, A
29	831.5	34.0	485	US-10-369-493-11863	Sequence 11863, A
30	831	34.0	458	US-10-369-493-10934	Sequence 10934, A
31	830	33.9	492	US-10-282-122A-49742	Sequence 49742, A
32	820	32.9	492	US-10-282-122A-49742	Sequence 49742, A
33	820	32.9	492	US-10-282-122A-49742	Sequence 49742, A
34	824	33.7	478	US-10-369-493-17031	Sequence 17031, A
35	822	33.6	478	US-10-369-493-17031	Sequence 17031, A
36	821.5	33.5	479	US-10-369-493-7523	Sequence 7523, Ap
37	821.5	33.5	481	US-10-369-493-4763	Sequence 4763, Ap
38	821.5	33.5	486	US-10-282-122A-49742	Sequence 49742, A
39	816	33.4	477	US-10-369-493-21219	Sequence 21219, A
40	815	33.3	537	US-10-424-599-162810	Sequence 162810, A
41	814.5	33.3	551	US-10-417-063-125211	Sequence 125211, A
42	813.5	33.3	482	US-10-369-493-8449	Sequence 8449, Ap
43	813.5	33.3	485	US-10-369-493-12251	Sequence 12251, A
44	813.5	33.3	489	US-10-282-122A-50658	Sequence 50658, A
45	812	33.2	460	US-10-369-493-16681	Sequence 16681, A

ALIGNMENTS

RESULT 1

US-09-830-751-6

Sequence 6, Application US/09830751

GENEPAI, INFORMATION:

APPLICANT: Suthers, Douglas F.

APPLICANT: Cameron, Douglas F.

TITLE OF INVENTION: Production of 3-Hydroxypropionic Acid in Recombinant

FILE REFERENCE: 2002/06, 2001/7

CURRENT APPLICATION NUMBER: US/09-830,751

PRIOR FILING DATE: 2000-08-30

PRIOR APPLICATION NUMBER: 60/151,440

PRIOR FILING DATE: 1999-08-30

PRIOR APPLICATION NUMBER: PCT/US00/23878

NUMBER OF SEQ ID NOS: 23

SOFTWARE: Patent In Ver. 2.1

SEQ ID NO: 6

LENGTH: 479

TYPE: PRT

ORGANISM: Escherichia coli

US-09-830-751-6

Query Match

Best Local Similarity 100.0%; Score 2446; DB 5; Length 479;

Matches 479; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY	1	MSVPVGHMPTITGQFVITWKA	11	VNPATFAVISTKPGQATLAKKATIAAFRAPEW	60
DB	1	MSVPVGHMPTITGQFVITWKA	11	VNPATFAVISTKPGQATLAKKATIAAFRAPEW	60
QY	61	EALEFAEFASWLEFKISACIFEFASFTSLIVETQYFGLAEVFAVTA-VITYMAFWAR	120		
DB	61	EALEFAEFASWLEFKISACIFEFASFTSLIVETQYFGLAEVFAVTA-VITYMAFWAR	120		
QY	121	REGGEIISDRGRENILLFKRALGVTTGILFMPNPFELIARKKAPALLTGTITIKPSEF	180		
DB	121	REGGEIISDRGRENILLFKRALGVTTGILFMPNPFELIARKKAPALLTGTITIKPSEF	180		
QY	181	TRNNATAPAKIVDPGLIPGVNIVLGGELFVGOELANPNVAVMSMGVSAGFKIMAT	240		
DB	181	TRNNATAPAKIVDPGLIPGVNIVLGGELFVGOELANPNVAVMSMGVSAGFKIMAT	240		
QY	241	AAFNITKYVTELEGRKAPAVMTATCTAVNAVIVSVYNSQVTRAFEEVYAKITPDQ	300		
DB	241	AAFNITKYVTELEGRKAPAVMTATCTAVNAVIVSVYNSQVTRAFEEVYAKITPDQ	300		
QY	301	FAVNELEGMQAVQFNPAFNPNTIAMGTLINAALEFVQKAPAVVERCAVAVPQAAVER	360		
DB	301	FAVNELEGMQAVQFNPAFNPNTIAMGTLINAALEFVQKAPAVVERCAVAVPQAAVER	360		

Db 401 FVNHGCEAMQAVQPCNIAERNDIAMGTLINAAALEREQKVAFAVEEAFVAGKAVEG 360
 QY 461 KGYVPTLLDVRQKMSIMHEETPGVLPVVAFTLEDAISMANISDYGITSSYTONL 420
 Db 461 KGYVPTLLDVRQKMSIMHEETPGVLPVVAFTLEDAISMANISDYGITSSYTONL 420
 QY 421 NVAMKAIKGIKFGETYINRENFAMQGFHAGKRSKSGIGADGKHGILGYLOTUVVYLOS 479
 Db 421 NVAMKAIKGIKFGETYINRENFAMQGFHAGKRSKSGIGADGKHGILGYLOTUVVYLOS 479

RESULT 2

US-10-469-493-823
 ? Sequence 823, Application US/10469493
 ? GENERAL INFORMATION:
 ? APPLICANT: Cao, Yompei
 ? APPLICANT: Hiroko, Gregory J.
 ? APPLICANT: Slater, Steven G.
 ? APPLICANT: Goldman, Barry S.
 ? TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ? FILE REFERENCE: 48-10(52052)H
 ? CURRENT APPLICATION NUMBER: US/10469,493
 ? PRIOR FILING DATE: 2003-02-28
 ? PRIOR APPLICATION NUMBER: US 60/460,039
 ? PRIOR FILING DATE: 2002-02-21
 ? NUMBER OF SEQ ID NOS: 47374
 ? SEQ ID NO 823
 ? LENGTH: 479
 ? TYPE: PRT
 ? ORGANISM: Escherichia coli
 US-10-469-493-823

Query Match 99.4%; Score 2432; Db 6; Length 479;
 Best Local Similarity 99.6%; Pred. No. 3,26-204;
 Matches 477; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MSVYVQHPKTIIGQVPTWKGDMIIIVNPAFAVISRIPDGAEDARKAIDAERAQDEW 60
 Db 1 MSVYVQHPKTIIGQVPTWKGDMIIIVNPAFAVISRIPDGAEDARKAIDAERAQDEW 60
 QY 61 EALPAIFRASMLRKISAGIRERASELSALIVEEGKIQOLAVEVAFTADYIDYMAEWA 120
 Db 61 EALPAIFRASMLRKISAGIRERASELSALIVEEGKIQOLAVEVAFTADYIDYMAEWA 120
 QY 121 RYEGEIIQSDRPGENITLKKRALGVTTGIIIPNPFPELLARKMAPALLTGNTVIKPEEF 180
 Db 121 RYEGEIIQSDRPGENITLKKRALGVTTGIIIPNPFPELLARKMAPALLTGNTVIKPEEF 180
 QY 181 TTNNALAPAKTIVEIQLPGGVFNLYVGRGETVGOELAGNPKYAMSGTSVAGCKIMAT 240
 Db 181 TTNNALAPAKTIVEIQLPGGVFNLYVGRGETVGOELAGNPKYAMSGTSVAGCKIMAT 240
 QY 241 AAKNITKVCLELGKAPATVMDADILEAVKATVDSRVINSQVNCFAERYVVGKGYDQ 300
 Db 241 AAKNITKVCLELGKAPATVMDADILEAVKATVDSRVINSQVNCFAERYVVGKGYDQ 300
 QY 401 FVNRLEFAMQAVQPCNIAERNDIAMGTLINAAALEREQKVAFAVEEAFVAGKAVEG 360
 Db 401 FVNRLEFAMQAVQPCNIAERNDIAMGTLINAAALEREQKVAFAVEEAFVAGKAVEG 360
 QY 461 KGYVPTLLDVRQKMSIMHEETPGVLPVVAFTLEDAISMANISDYGITSSYTONL 420
 Db 461 KGYVPTLLDVRQKMSIMHEETPGVLPVVAFTLEDAISMANISDYGITSSYTONL 420
 QY 421 NVAMKAIKGIKFGETYINRENFAMQGFHAGKRSKSGIGADGKHGILGYLOTUVVYLOS 479
 Db 421 NVAMKAIKGIKFGETYINRENFAMQGFHAGKRSKSGIGADGKHGILGYLOTUVVYLOS 479

RESULT 3
 US-10-446-203-14820

? Sequence 13820, Application US/10446203
 ? GENERAL INFORMATION:
 ? APPLICANT: Gary Breton et al
 ? TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ? FILE REFERENCE: 2709, 2004001
 ? CURRENT APPLICATION NUMBER: US/10446,203
 ? PRIOR FILING DATE: 2003-05-27
 ? PRIOR APPLICATION NUMBER: US/09/489,039
 ? PRIOR FILING DATE: 2000-01-27
 ? PRIOR APPLICATION NUMBER: US 6,0117,747
 ? PRIOR FILING DATE: 1999-01-29
 ? NUMBER OF SEQ ID NOS: 14342
 ? SEQ ID NO 13820
 ? LENGTH: 514
 ? TYPE: PRT
 ? ORGANISM: Klebsiella pneumoniae
 US-10-446-203-13820

Query Match 88.2%; Score 2157; Db 6; Length 514;
 Best Local Similarity 85.8%; Pred. No. 4,66-180;
 Matches 411; Conservative 43; Mismatches 25; Indels 0; Gaps 0;

QY 1 MSVYVQHPKTIIGQVPTWKGDMIIIVNPAFAVISRIPDGAEDARKAIDAERAQDEW 60
 Db 36 MTAPOVHPMYIDQFVSGKDMIIIVNPAFAVISRIPDGAEDARKAIDAERAQDEW 95
 QY 61 EALPAIFRASMLRKISAGIRERASELSALIVEEGKIQOLAVEVAFTADYIDYMAEWA 120
 Db 46 EALPAIFRASMLRKISAGIRERASELSALIVEEGKIQOLAVEVAFTADYIDYMAEWA 155
 QY 121 RYEGEIIQSDRPGENITLKKRALGVTTGIIIPNPFPELLARKMAPALLTGNTVIKPEEF 180
 Db 156 RYEGEIIQSDRPGENITLKKRALGVTTGIIIPNPFPELLARKMAPALLTGNTVIKPEEF 215
 QY 181 TTNNALAPAKTIVEIQLPGGVFNLYVGRGETVGOELAGNPKYAMSGTSVAGCKIMAT 240
 Db 216 TTNNALAPAKTIVEIQLPGGVFNLYVGRGETVGOELAGNPKYAMSGTSVAGCKIMAT 275
 QY 241 AAKNITKVCLELGKAPATVMDADILEAVKATVDSRVINSQVNCFAERYVVGKGYDQ 300
 Db 276 AAKNITKVCLELGKAPATVMDADILEAVKATVDSRVINSQVNCFAERYVVGKGYDQ 335
 QY 301 FVNRLEFAMQAVQPCNIAERNDIAMGTLINAAALEREQKVAFAVEEAFVAGKAVEG 360
 Db 336 FVNRLEFAMQAVQPCNIAERNDIAMGTLINAAALEREQKVAFAVEEAFVAGKAVEG 395
 QY 461 KGYVPTLLDVRQKMSIMHEETPGVLPVVAFTLEDAISMANISDYGITSSYTONL 420
 Db 396 KGYVPTLLDVRQKMSIMHEETPGVLPVVAFTLEDAISMANISDYGITSSYTONL 455
 QY 421 NVAMKAIKGIKFGETYINRENFAMQGFHAGKRSKSGIGADGKHGILGYLOTUVVYLOS 479
 Db 456 NVAMKAIKGIKFGETYINRENFAMQGFHAGKRSKSGIGADGKHGILGYLOTUVVYLOS 514

RESULT 4

US-10-282-122A-65716

? Sequence 65716, Application US/10282122A
 ? GENERAL INFORMATION:
 ? APPLICANT: Wang, Jiansu
 ? APPLICANT: Zamudio, Carlos
 ? APPLICANT: Malone, Cheryl
 ? APPLICANT: Haselbeck, Robert
 ? APPLICANT: Ohlsen, Kari
 ? APPLICANT: Zyskind, Judith
 ? APPLICANT: Wall, Daniel
 ? APPLICANT: Trawick, John
 ? APPLICANT: Carr, Grant
 ? APPLICANT: Yamamoto, Robert
 ? APPLICANT: Forsyth, R.
 ? APPLICANT: Xu, H.
 ? TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

INTERNET: 1997-00-14

1. **Introduction**
 2. **Background**
 3. **Methods**
 4. **Results**
 5. **Conclusion**
 6. **References**
 7. **Appendix**
 8. **Figure 1**
 9. **Figure 2**
 10. **Figure 3**
 11. **Figure 4**
 12. **Figure 5**
 13. **Figure 6**
 14. **Figure 7**
 15. **Figure 8**
 16. **Figure 9**
 17. **Figure 10**
 18. **Figure 11**
 19. **Figure 12**
 20. **Figure 13**
 21. **Figure 14**
 22. **Figure 15**
 23. **Figure 16**
 24. **Figure 17**
 25. **Figure 18**
 26. **Figure 19**
 27. **Figure 20**
 28. **Figure 21**
 29. **Figure 22**
 30. **Figure 23**
 31. **Figure 24**
 32. **Figure 25**
 33. **Figure 26**
 34. **Figure 27**
 35. **Figure 28**
 36. **Figure 29**
 37. **Figure 30**
 38. **Figure 31**
 39. **Figure 32**
 40. **Figure 33**
 41. **Figure 34**
 42. **Figure 35**
 43. **Figure 36**
 44. **Figure 37**
 45. **Figure 38**
 46. **Figure 39**
 47. **Figure 40**
 48. **Figure 41**
 49. **Figure 42**
 50. **Figure 43**
 51. **Figure 44**
 52. **Figure 45**
 53. **Figure 46**
 54. **Figure 47**
 55. **Figure 48**
 56. **Figure 49**
 57. **Figure 50**
 58. **Figure 51**
 59. **Figure 52**
 60. **Figure 53**
 61. **Figure 54**
 62. **Figure 55**
 63. **Figure 56**
 64. **Figure 57**
 65. **Figure 58**
 66. **Figure 59**
 67. **Figure 60**
 68. **Figure 61**
 69. **Figure 62**
 70. **Figure 63**
 71. **Figure 64**
 72. **Figure 65**
 73. **Figure 66**
 74. **Figure 67**
 75. **Figure 68**
 76. **Figure 69**
 77. **Figure 70**
 78. **Figure 71**
 79. **Figure 72**
 80. **Figure 73**
 81. **Figure 74**
 82. **Figure 75**
 83. **Figure 76**
 84. **Figure 77**
 85. **Figure 78**
 86. **Figure 79**
 87. **Figure 80**
 88. **Figure 81**
 89. **Figure 82**
 90. **Figure 83**
 91. **Figure 84**
 92. **Figure 85**
 93. **Figure 86**
 94. **Figure 87**
 95. **Figure 88**
 96. **Figure 89**
 97. **Figure 90**
 98. **Figure 91**
 99. **Figure 92**
 100. **Figure 93**
 101. **Figure 94**
 102. **Figure 95**
 103. **Figure 96**
 104. **Figure 97**
 105. **Figure 98**
 106. **Figure 99**
 107. **Figure 100**
 108. **Figure 101**
 109. **Figure 102**
 110. **Figure 103**
 111. **Figure 104**
 112. **Figure 105**
 113. **Figure 106**
 114. **Figure 107**
 115. **Figure 108**
 116. **Figure 109**
 117. **Figure 110**
 118. **Figure 111**
 119. **Figure 112**
 120. **Figure 113**
 121. **Figure 114**
 122. **Figure 115**
 123. **Figure 116**
 124. **Figure 117**
 125. **Figure 118**
 126. **Figure 119**
 127. **Figure 120**
 128. **Figure 121**
 129. **Figure 122**
 130. **Figure 123**
 131. **Figure 124**
 132. **Figure 125**
 133. **Figure 126**
 134. **Figure 127**
 135. **Figure 128**
 136. **Figure 129**
 137. **Figure 130**
 138. **Figure 131**
 139. **Figure 132**
 140. **Figure 133**
 141. **Figure 134**
 142. **Figure 135**
 143. **Figure 136**
 144. **Figure 137**
 145. **Figure 138**
 146. **Figure 139**
 147. **Figure 140**
 148. **Figure 141**
 149. **Figure 142**
 150. **Figure 143**
 151. **Figure 144**
 152. **Figure 145**
 153. **Figure 146**
 154. **Figure 147**
 155. **Figure 148**
 156. **Figure 149**
 157. **Figure 150**
 158. **Figure 151**
 159. **Figure 152**
 160. **Figure 153**
 161. **Figure 154**
 162. **Figure 155**
 163. **Figure 156**
 164. **Figure 157**
 165. **Figure 158**
 166. **Figure 159**
 167. **Figure 160**
 168. **Figure 161**
 169. **Figure 162**
 170. **Figure 163**
 171. **Figure 164**
 172. **Figure 165**
 173. **Figure 166**
 174. **Figure 167**
 175. **Figure 168**
 176. **Figure 169**
 177. **Figure 170**
 178. **Figure 171**
 179. **Figure 172**
 180. **Figure 173**
 181. **Figure 174**
 182. **Figure 175**
 183. **Figure 176**
 184. **Figure 177**
 185. **Figure 178**
 186. **Figure 179**
 187. **Figure 180**
 188. **Figure 181**
 189. **Figure 182**
 190. **Figure 183**
 191. **Figure 184**
 192. **Figure 185**
 193. **Figure 186**
 194. **Figure 187**
 195. **Figure 188**
 196. **Figure 189**
 197. **Figure 190**
 198. **Figure 191**
 199. **Figure 192**
 200. **Figure 193**
 201. **Figure 194**
 202. **Figure 195**
 203. **Figure 196**
 204. **Figure 197**
 205. **Figure 198**
 206. **Figure 199**
 207. **Figure 200**
 208. **Figure 201**
 209. **Figure 202**
 210. **Figure 203**
 211. **Figure 204**
 212. **Figure 205**
 213. **Figure 206**
 214. **Figure 207**
 215. **Figure 208**
 216. **Figure 209**
 217. **Figure 210**

```
QY      9 MYIDGQZVWTRCNDAMINVNPNATEAVISPIPDQAFAKPAIIIAAFVAQPEWEALPAIER 68
```

```

Db      5  LITDGMWSSSGKTVKSPVIGVIGKFAATRHIVKRIIAAFNAFMANNDLNSVER 64
      69  ASMLRKISAGIPEPASEISALIVEGKIQGLAEVNAFTADYIDYMAEMARKEVEILQ 128
      65  SKIIFYAKELIEKNAPLENIIMEENKPVKEAKEVEVIGVILQVUYAEMARKLNGEVE 124
OY      129  SDPGEENILFKRALGVTTGILPWNFPFLIARKKAPALITGNTIVIKPSEFTTNNAIAF 188
      125  GTSSHHKIFQYKVPYGVIVALLPMPFPCMAVARKIAPALLGNTIVLKPSSDPGSAEMI 184
      189  AKIVDPGIGPCVFLVIGPCEVYQGLACNPKVAMVSMTSVSAEKKIMATAKNTKV 248
      185  VKRFEAGVPRKVLNFTIGRGEICDYIVEHKKNVLLITMGSTATGDIRMOKASANAKL 244
OY      249  CLEGGKAPAIVMDADLELAVKAIIVSRVINSQVNCARVYVQKIGYDQFVNRUGEA 308
      245  IELGOKAFEFYWKADMDNMLKTLIMAKVYNACOSCIABERLYVHDYIDTFMSREVEL 304
OY      303  MAAYVGNPAEKNDIAMKPLINAAALKEVAVARAVEGAKVAPGCK--AVEG--KGY 363
      305  SKRLLDGPRKND--MGPLINKGALQNTSEIVEPAKFSKAKILFGSGQPSLQSPYNGY 361
OY      364  YVPPTLLDVRGEMSMHEETGPAVPAFTLEDAISMANSDYGTSSITYONLVA 423
      362  FLPLTIGNADOKSKIPOEFTFAPYIGARKISSVEEMVDJANDSKYGLASTYLFTRKNTI 421
OY      424  MKAIGLAFGEIYINRENFEMOGFHAGKRSKIGGADGKHGLHGYLOTOVYL 477
      422  PEASERIRPGELVYMMGCFEASOGYHTGFRMTGOAGGCSKYCISEYILKKNIV 475
Db

```

RESULT 9

Sequence 69547, Application US/10282122A

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu

APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari

APPLICANT: Zyskind, Judith

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John

APPLICANT: Carr, Grant

APPLICANT: Yamamoto, Robert

APPLICANT: Forsyth, P.

APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-04-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-24

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-06

PRIOR APPLICATION NUMBER: 60/269,408

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See file wrapper or PAM

NUMBER OF SEQ ID NOS: 78614

```

: SOFTWARE: PatentIn version 3.1
: SEQ ID NO: 69547
: LENGTH: 480
: TYPE: PR
: ORGANISM: Pseudomonas syringae
US-10-282-122A-69547

Query Match      30.8%; Score 900.5; Ds 6; Length 480;
Best Local Similarity 39.3%; Pred. No. 5,36-70;
Matches 184; Conservative 101; Mismatches 180; Indels 3; Gaps 2;

OY      16  YTDGZPVIMVQAMIVVNIATEAVISPTFFZAMETAKKITAFAFQPPWFLPAIFA 69
      14  YINQJMLIAHISQSKIKVNNPATINELISTVPMGAETBRAIEADKALPMAWLTAKFPG 73
OY      70  SWIFKISAGIPEPASEISALIVEGKIQGLAEVNAFTADYIDYMAEMARKEVEILQ 129
      74  NKLPRWFLMELNLELDLMLILEGKLAARKEIITYASPIEMFAEKRPVYGVIG 133
OY      130  DRPGFNILFKRALGVTTGILPWNFPFLIARKKAPALITGNTIVIKPSEFTTNNAIAF 189
      134  HOPDKRLIVLKOPIGVTAITPWNFPAMITRKGPALAGCTMVLPASQTFPSALALA 193
OY      190  KIVDEIGLPRGVFNVLUGSEIVGQELAGNPKVAMVSMTSVSAEKKIMATAKNTKV 249
      194  ELAEKAGIPAGVFSVVISAGDISLIGNIVKLSSTISIEIGKQLMARCKADIKKYS 253
OY      250  EELGKAPAIVMDADLELAVKAIIVSRVINSQVNCARVYVQKIGYDQFVNRUGEA 309
      254  TELGNAPEIYFDADLDKAVEGAMISKYRNNGTCVCAIRIYQDDGYDAFAEKLKVV 313
OY      310  QAVQGNPAERNDIAMKPLINAAALKEVAVARAVEGAKVAPGKAYVPPPTL 369
      314  GKELKISGDE DGLIGLIEKAVAKKIMIAVAKGAVLIGSGLCS--FLIL 376
OY      370  LLDVRGEMSMHEETGPAVPAFTLEDAISMANSDYGTSSITYONLVAAMAIG 429
      371  LVNSKDAVAAREEFPGLAPLFPRKEAEKIALALANTFPEIASYFAVAGMSVFPVAEA 430
OY      430  LKFGPIYINRENFEMOGFHAGKRSKIGGADGKHGLHGYLOTOVYL 477
      431  LEYGVGINTGLISNELAPFGGKSSGLGREGSKYGEIDYLEIKYDL 478
Db

```

RESULT 10

Sequence 70076, Application US/10282122A

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu

APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari

APPLICANT: Zyskind, Judith

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John

APPLICANT: Carr, Grant

APPLICANT: Yamamoto, Robert

APPLICANT: Forsyth, R.

APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-04-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

RESULT 11
 US 10 469 493-13571
 1 Sequence, 13571, Application US/10469493
 2 GENERAL INFORMATION:
 3 APPLICANT: Cao, Yongwei
 4 APPLICANT: Hinkley, Gregory J.
 5 APPLICANT: Slater, Steven C.
 6 APPLICANT: Goldman, Barry S.
 7 APPLICANT: Chen, Xiantong
 8 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 9 TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 10 FILE REFERENCE: 39-10(52052)B
 11 CURRENT APPLICATION NUMBER: US/10/369,493
 12 CURRENT FILING DATE: 2003-02-28
 13 PRIOR APPLICATION NUMBER: US 60/360,039
 14 PRIOR FILING DATE: 2002-02-21

```

RESULT 12
US-10-282-122A-43455
Sequence 43455, Application US/10282122A
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA-034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/267,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,435
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/240,447

```

```

; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PAM.
; SOFTWARE: Patent version 3.1
; SEQ ID NO 43455
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-43455

```

```

Query Match          36.0%; Score 879.5; DB 6; Length 483;
Best Local Similarity 39.3%; Pred. No. 3,7e+68;
Matches 184; Conservative 96; Mismatches 187; Indels 1; Gaps 1;

```

```

QY 10 YIDGCFVTWRGDAMIDVVPATEAVISRIIPDGAEDARKAIDAERAPPEMALPATEPA 69
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 14 YVDCAMVADNDNGCITKVNPPATIGETISVPMKGAETRALEADKALPANKALIAKERA 73
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 70 SWLRKISAGIRERASEISALIVEEGKIQQLAEVEVATADYIDYMAEMARPEGETTQS 129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 74 NKLRFWFLMIENQDLARIMTEGSKPLAAKSHIAVAASFLEMPGEAKRIYGRIPDS 133
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 130 DDPGNNILIFKPAICVTGIIIPWNPFFLIAPKAPALITGNTIVTSPSETTNATAPA 189
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 134 HCPDKRTIVIKQPIGVTAITTFWNPFSAMITPKAPALAACTWVLKPASTPYSAIALA 194
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 190 KIVDEIGLPGVFNVLGRGETVGOELAGNPKVAMVMTGVSASGEKIMATAKNITKYC 249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 194 ELAEPAKIKVSVSVTSASAEVSEELISNPVTKLFTSTELGPDLMACQADIKKVS 254
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 250 LELGKAPAIYMDADLEAVKAIYDSRVINSQVWVWVAVVYVYKSIYIGFVVKLGEM 309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 254 LELGNAPEIYEDDADLAAVEGALISKYPNNQGVCANLYQDVYDAFVDKRAAV 424
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 310 QAVGQPAERNDIMGSPLIINAALFVEYKVAFAVEGSAVAFGSKAVEGKYYPPTL 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 314 AKLINSNGLEAG-VITGPIIAKAKAVAEHIALAVSKAKAVSSKKHALGTFEPTI 483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 370 LITVQEMSIHMETFGPVLPVAVFTLEPAISMANSTVYCTSTYTQNTNVAMKATKS 429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 373 LVDVPRKNAIVSKDETFGLAPVFRKLEAEVIAVMSNTERGLASVFAKHLAVFRVAEQ 442
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 430 LKPGETVINKENFEAMUGFHAMKRSKSGVADKSHGLGYIGTQVYVL 477
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 433 LEYGMVINTGILISNEVAPFGSIKASJGPRGSKYIGEDYLEIKYICL 491
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 13
US-10-366-683-27749
; Sequence 27749, Application US/10366683
; GENERAL INFORMATION:
; APPLICANT: Rubenfeld, Marc J.
; APPLICANT: Nolling, Jork
; APPLICANT: Deloughery, Craig
; APPLICANT: Bush, David
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PAT03-04
; CURRENT APPLICATION NUMBER: US/10/366,683
; PRIOR FILING DATE: 2003-02-13
; PRIOR APPLICATION NUMBER: 03/252,991
; PRIOR FILING DATE: 1999-02-18
; NUMBER OF SEQ ID NOS: 33142

```

```

; SEQ ID NO 27749
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-366-683-27749

```

```

Query Match          36.0%; Score 879.5; DB 6; Length 594;
Best Local Similarity 39.3%; Pred. No. 5,2e+68;
Matches 184; Conservative 96; Mismatches 187; Indels 1; Gaps 1;

```

```

QY 10 YIDGCFVTWRGDAMIDVVPATEAVISRIIPDGAEDARKAIDAERAPPEMALPATEPA 69
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 125 YVDCAMVADNDNGCITKVNPPATIGETISVPMKGAETRALEADKALPANKALIAKERA 184
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 70 SWLRKISAGIRERASEISALIVEEGKIQQLAEVEVATADYIDYMAEMARPEGETTQS 129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 185 NKLRFWFLMIENQDLARIMTEGSKPLAAKSHIAVAASFLEMPGEAKRIYGRIPDS 244
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 130 DDPGNNILIFKPAICVTGIIIPWNPFFLIAPKAPALITGNTIVTSPSETTNATAPA 189
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 245 HCPDKRTIVIKQPIGVTAITTFWNPFSAMITPKAPALAACTWVLKPASTPYSAIALA 304
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 190 KIVDEIGLPGVFNVLGRGETVGOELAGNPKVAMVMTGVSASGEKIMATAKNITKYC 249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 305 ELAEPAKIKVSVSVTSASAEVSEELISNPVTKLFTSTELGPDLMACQADIKKVS 364
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 250 LELGKAPAIYMDADLEAVKAIYDSRVINSQVWVWVAVVYVYKSIYIGFVVKLGEM 309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 365 LELGNAPEIYEDDADLAAVEGALISKYPNNQGVCANLYQDVYDAFVDKRAAV 424
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 310 QAVGQPAERNDIMGSPLIINAALFVEYKVAFAVEGSAVAFGSKAVEGKYYPPTL 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 425 AKLINSNGLEAG-VITGPIIAKAKAVAEHIALAVSKAKAVSSKKHALGTFEPTI 483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 370 LITVQEMSIHMETFGPVLPVAVFTLEPAISMANSTVYCTSTYTQNTNVAMKATKS 429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 484 LVDVPRKNAIVSKDETFGLAPVFRKLEAEVIAVMSNTERGLASVFAKHLAVFRVAEQ 543
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 430 LKPGETVINKENFEAMUGFHAMKRSKSGVADKSHGLGYIGTQVYVL 477
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 544 LEYGMVINTGILISNEVAPFGSIKASJGPRGSKYIGEDYLEIKYICL 591
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 14
US-10-419-128-27749
; Sequence 27749, Application US/10419128
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfeld et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196,136
; CURRENT APPLICATION NUMBER: US/10/419,128
; PRIOR FILING DATE: 2003-04-21
; PRIOR APPLICATION NUMBER: US/09/752,391
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27749
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-419-128-27749

```

```

Query Match          36.0%; Score 879.5; DB 6; Length 594;
Best Local Similarity 39.3%; Pred. No. 5,2e+68;
Matches 184; Conservative 96; Mismatches 187; Indels 1; Gaps 1;

```

```

QY 70 SMLEKTSAGLREKASLALVEEKKLQGLAEVAVATADYLDYMAWARRRDEETIGS 129
DB 185 NKLRRKFDLMLENDOLARLMTLEOKPLAEKGEIATVASELEMGEEFAKPIYGGTITG 244
QY 130 DREGENILLERKALGVTTGILPNNPPEPLIARRMAVALTGNTIVIKPSEFTTNNALFA 189
DB 245 HOPKRLIIVIKOPIGVTAITPNNPFSAMITRRKAGALAAAGTMVLKPSQTFPSALALA 304
QY 190 KIVETIGLPGGVFNILVLRGETVGOELAGNPKYAMVSMTGSVSAEKKIMATAAKNTTKYC 249
DB 405 FLARRACIIPRGVSVVTVGSAGEGGELTSMIVKRLITGSLIPIGQIMAECAODIKRAS 364
QY 250 LELGKAPAIYMDADLELAVKAIIVDSRVINSQGVNCAERVYVGKGTIDQVNRIGEM 309
DB 465 LELGKAPAIYVDDADLELAVKAIIVDSRVINSQGVNCAERVYVGKGTIDQVNRIGEM 424
QY 410 QAVQGNPARRNDIAMGRLINAAALFRVEOKVARAAVEGARAAGKAVEGKGYPPPTL 369
DB 425 AKINIGNGLEAG-VTTGFLIDAKAVAKVEEHIDAAVSKAGAKVSGKPHALGTPEPTI 483
QY 470 LIAVROBMSIMHEETFGVLPYVAFDTLEDAISMANDSDYGLTSSITYONTLVAMKATKG 429
DB 484 LIAVPRKALVSKDPTFGPLAPVRFKDEADVIAKMSNDTPEGLASTFYAADLARVRVAVQ 543
QY 430 LKPGETYINRENPEAMOGFHAWRKSGTIGADKRGHGLHYIQTQVYL 477
DB 544 LEYGVGCIINTGLISNEVAPFGGIKASGLGHSKCYGIEIYILEIKYLCI 591

```

RESULT 15

US-10-282-122A-56650

Sequence: 56650, Application US/10282122A

GENERAL INFORMATION:

```

APPLICANT: Mamu, Hiansu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsson, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
PRIOR FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-04-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-24
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/210,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/210,447
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-10
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: Patent in version 3.1

```

```

: SEQ ID NO 56650
: LENGTH: 482
: TYPE: PRF
: ORGANISM: Escherichia coli
US-10-282 122A-56650
Query Match 35.7%; Score 873.5; DB 6; Length 482;
Best Local Similarity 39.5%; Pred. No. 1.2e+67;
Matches 182; Conservative 91; Mismatches 187; Indels 1; Gaps 1;
11 IDGQFVTKGDAVIDVNPATEAVISRIHQQAELAKKAIIDAERKADPEWALPAIERAS 70
15 INGEMLDANNGEAIDVTNPANGIDKLGSSVPAKADFTRAIDANBALPAMRALTKERAT 74
71 WLRKTSAGTPEPASFSIALIVPEGSKTQGLAVFVAFVFTNDYLDYMAWARRRDEETIGS 130
75 ILRNWFLMKEHODDLARLMTLEOKPLAEKGEIATVASELEMGEEFAKRIYGGTITGH 134
131 KPGHLLERKALGVTTGILPNNPPEPLIARRMAVALTGNTIVIKPSEFTTNNALFAAK 190
135 QADKRLIIVIKOPIGVTAITPNNPFSAMITRRKAGPALAGTMVLKPSQTFPSALALAE 194
191 IVNFTGLPGRVFNILVLRGETVGOELAGNPKYAMVSMTGSVSAEKKIMATAAKNTTKY 250
195 LAIRAGVAGVFNVVTVGSAGAVGNPLTSMPIVYKLSFTSTETIGQIMAECAODIKRAS 254
251 LELGKAPAIYMDADLELAVKAIIVDSRVINSQGVNCAERVYVGKGTIDQVNRIGEM 310
255 ELGQNAFIVFDADLELAVKAIIVDSRVINSQGVNCAERVYVGKGTIDQVNRIGEM 314
311 AVQGNPARRNDIAMGRLINAAALFRVEOKVARAAVEGARAAGKAVEGKGYPPPTL 370
315 KLHIGDGLD-NGVTIGPLDEKAVAKVEEHIDALEKAGAVVGGKAHERGNGFPPTL 373
371 LDVROBMSIMHEETFGVLPYVAFDTLEDAISMANDSDYGLTSSITYONTLVAMKATKG 430
374 VDVPAKAKSKKEETFGPLAPVRFKDEADVIAKMSNDTPEGLASTFYAADLARVRVAVQ 433
431 KPGETYINRENPEAMOGFHAWRKSGTIGADKRGHGLHYIQTQVYL 471
434 FYGVGCIINTGLISNEVAPFGGIKASGLGHSKCYGIEIYILE 474

```

Search completed: June 24, 2003, 10:35:22
Job time: 52.8555 secs

GenBank version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Jan 24, 2003, 10:10:05, Search time 18.3303 Seconds

(without alignments)
2512 144 Million cell updates/sec

Title: US-09-830-751-6

Perfect score: 2446

Sequence: 1 MSVVVGHMPTIDGQFVITWIS ... ALGKHSLSHYLTGVVYVLS 474

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 100%

Database :
1: PIR.73.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARY

Result No.	Score	Query Match	Length	DB ID	Description
1	2432	99.4	479	2 A38165	lactaldehyde dehyd
2	2426	99.2	479	2 E00881	aldehyde dehydrog
3	2426	99.2	479	2 D85737	aldehyde dehydrog
4	1557	63.7	480	2 A81023	aldehyde dehydrog
5	1552	63.5	480	2 F81965	lactaldehyde dehyd
6	1279.5	52.3	393	2 F81394	probable lactaldeh
7	879.5	36.0	483	2 F81394	succinate-semialde
8	879.5	35.7	482	2 F65045	succinate-semialde
9	870.5	35.6	482	2 B91069	succinate-semialde
10	870.5	35.6	482	2 D85213	succinate-semialde
11	858.5	35.1	482	2 A80844	succinate-semialde
12	845.5	34.6	475	2 D81064	succinate-semialde
13	836	34.2	491	2 C95248	probable succinate
14	835	34.1	468	2 C83774	succinate-semialde
15	822	33.6	482	2 F81648	succinate-semialde
16	821.5	33.6	487	2 A13300	succinate-semialde
17	814.5	33.3	484	2 A95389	succinate-semialde
18	814.5	33.3	484	2 E96825	hypothetical prote
19	811.5	33.2	488	2 A11188	succinate-semialde
20	809	33.1	488	2 A11546	succinate-semialde
21	804.5	32.9	476	2 A79643	glycine betaine al
22	804	32.9	495	2 A75608	aldehyde dehydrog
23	804	32.9	472	2 C94774	succinate-semialde
24	798.5	32.6	486	2 A98308	actK protein (0594
25	798.5	32.6	486	2 A82975	succinate-semialde
26	795.5	32.5	484	2 A83176	NAD-dependent succ
27	795.5	32.5	490	2 E95316	succinate-semialde
28	795	32.5	494	2 B95411	probable aldehyde
29	791.5	32.4	487	2 C96279	CaM3 succinate-se

RESULT 1

A38165

lactaldehyde dehydrogenase (EC 1.2.1.22) alda - Escherichia coli (strain K-12)

C.Dat: 28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change 01-Mar-2002

C.Species: Escherichia coli

C.Accession: A38165; B64893

R.Hidalgo, E.; Chen, Y.M.; Lin, E.C.C.; Aguilar, J.

J. Bacteriol. 173, 6118-6123, 1991

A.Title: Molecular cloning and DNA sequencing of the Escherichia coli K-12 ald gene

A.Reference number: A38165; MIM 92011371; PMID 1917845

A.Accession: A38165

A.Status: preliminary

A.Molecule type: DNA

A.Residues: 1-479 <HID>

A.Cross-references: GR M64511, NID:J15221, PIR:AA023427, PDB:J15222

R.Hidalgo, E.; Chen, Y.M.; Lin, E.C.C.; Aguilar, J.

J. Bacteriol. 173, 6118-6123, 1991

A.Title: Molecular cloning and DNA sequencing of the Escherichia coli K-12 ald gene

A.Reference number: A38165; MIM 92011371; PMID 1917845

A.Accession: A38165

A.Status: preliminary

A.Molecule type: DNA

A.Residues: 1-479 <HID>

A.Cross-references: GR M64511, NID:J15221, PIR:AA023427, PDB:J15222

R.Hidalgo, E.; Chen, Y.M.; Lin, E.C.C.; Aguilar, J.

J. Bacteriol. 173, 6118-6123, 1991

A.Title: Molecular cloning and DNA sequencing of the Escherichia coli K-12 ald gene

A.Reference number: A38165; MIM 92011371; PMID 1917845

ALIGNMENTS

Db	121	IVYSHIIQSDPPGPNILIFKRAVIVITGIIIPWNPFFLTARKKAPALITGNTIVIKPSPF	180
QY	121	IVYSHIIQSDPPGPNILIFKRAVIVITGIIIPWNPFFLTARKKAPALITGNTIVIKPSPF	180
Db	61	FAIIAIFPASMUKISAW::FFKASPIAALLVDSKIKGIAVYVAFATAYIDVAPVAP	120
QY	61	FAIIAIFPASMUKISAW::FFKASPIAALLVDSKIKGIAVYVAFATAYIDVAPVAP	120
Db	1	MSVVVGHMPTIDGQFVITWIS	60
QY	1	MSVVVGHMPTIDGQFVITWIS	60
QY	99.4%	Score 2432	DB ID Length 479
QY	Best Local Similarity	99.4%	Pred. No. 5.4e-122
QY	Matches	477	Conservative 0; Mismatches 2; Indels 0; Gaps 0;
A:Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-479 <HID>			
A:Cross-references: GR:M4541, NID:345221, P10N:AA023427.1, P1D:q145222			
R:Rattler, F.P.; Plunkett III, G.; Bloch, C.A.; Petina, N.T.; Burland, V.; Riley, M.			
Science 277, 1453-1462, 1997			
Article: The complete genome sequence of Escherichia coli K-12.			
A:Reference number M4720, M1D:9726617, PMID:9278503			
A:Accession: B64893			
A:Status: nucleic acid sequence not shown; translation not shown			
A:Molecule type: DNA			
A:Residues: 1-479 <BLAT>			
A:Cross-references: GR:AE000239, GR:000096, NID:3187662, P10N:AA074497.1, P1D:q18766			
A:Experimental source, strain K-12, substrain MG1655			
C:Genetics:			
A:Gene: aldA			
C:Function:			
A:Description: catalyzes oxidation of lactaldehyde to lactate using NAD			
A:Note: acts also on other aldehydes			
C:Superfamily: aldehyde dehydrogenase (NMD+), aldehyde dehydrogenase homology			
C:Keywords: NAD, oxidoreductase			
E:44-305/Domain, aldehyde dehydrogenase [EC:1.1.1.41] M1D:			
F:178-257/Domain, NAD binding *status predicted *NMD:			
F:251,285/Active site, Glu, Cys #status predicted			

181 TTNNATAFKATVDEIGLPRGVNVLVGRGFTVGOELAGNPKVAMVSMTGSAGKIMAT 240
 Db TPNNATAFKATVDEIGLPRGVNVLVGRGFTVGOELAGNPKVAMVSMTGSAGKIMAT 240
 QY 241 AAKNITKVCLEIGGKAPATVMDADDELAVKATVDSRVINSQVCCACARVYVVGITDQ 300
 Db 241 AAKNITKVCLEIGGKAPATVMDADDELAVKATVDSRVINSQVCCACARVYVVGITDQ 300
 QY 401 FVNRLEFAMQAVQFNGNPAERNDIAMGPIINAAALFEVEKQVAVAEAGARVAFGKAVEG 360
 Db 401 FVNRLEFAMQAVQFNGNPAERNDIAMGPIINAAALFEVEKQVAVAEAGARVAFGKAVEG 360
 QY 461 KGYVPTPTLLDVRQKMSIMHEETFGPVLPVVAFTLEDAISMANDSDYGLTSSITDQ 420
 Db 461 KGYVPTPTLLDVRQKMSIMHEETFGPVLPVVAFTLEDAISMANDSDYGLTSSITDQ 420
 QY 421 NVAMKAIKGLKFGFTYINRENFAMQGFHAGMKKSGTGGADCKHGLHFYLOTQVYVLOS 479
 Db 421 NVAMKAIKGLKFGFTYINRENFAMQGFHAGMKKSGTGGADCKHGLHFYLOTQVYVLOS 479

RESULT 2

Query Match 99.28; Score 2426; DB 2; Length 479;
 Best Local Similarity 99.48; Pred. No. 1,76-161;
 Matches 476; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 A:Species: *Escherichia coli*
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 63-Aug-2001
 C:Accession: E90881
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
 Yasunaga, N.; Yasunaga, T.; Kubota, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8: 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gene
 A:Reference number: A95629; MUID:21156231; PMID:11298796
 A:Accession: E90881
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1,479 <STO>
 A:Cross references: GB:BA000007; PDB:BAH3444.1; PDB:q1361487; GSPDB:GN00154
 A:Experimental source: strain 0157:H7, substrain RMD 0509952
 C:Genetics:
 A:Gene: EHEC201
 C:Superfamily: aldohyde dehydrogenase (NAD+); aldohyde dehydrogenase homology

Query Match 99.28; Score 2426; DB 2; Length 479;
 Best Local Similarity 99.48; Pred. No. 1,76-161;
 Matches 476; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MSVPVQHPMTDGGFTWGRDAMIIVVNPATFAVISRIIPDQAFARATDAEAPQEW 60
 Db 1 MSVPVQHPMTDGGFTWGRDAMIIVVNPATFAVISRIIPDQAFARATDAEAPQEW 60
 QY 61 EALPAIERASWMLKISAGIREBASISALIVEGKIQOLAEVVAFTADYIDYAEAR 120
 Db 61 EALPAIERASWMLKISAGIREBASISALIVEGKIQOLAEVVAFTADYIDYAEAR 120
 QY 121 RYGEETIOSDRGENILLFKRALGYTTGILPNPFEPFLARKMAFALLTGNTIVIKSEF 180
 Db 121 RYGEETIOSDRGENILLFKRALGYTTGILPNPFEPFLARKMAFALLTGNTIVIKSEF 180
 QY 181 KYEETIIOSDRGENILLFKRALGYTTGILPNPFEPFLARKMAFALLTGNTIVIKSEF 180
 Db 181 KYEETIIOSDRGENILLFKRALGYTTGILPNPFEPFLARKMAFALLTGNTIVIKSEF 180
 QY 241 AAKNITKVCLEIGGKAPATVMDADDELAVKATVDSRVINSQVCCACARVYVVGITDQ 300
 Db 241 AAKNITKVCLEIGGKAPATVMDADDELAVKATVDSRVINSQVCCACARVYVVGITDQ 300
 QY 401 FVNRLEFAMQAVQFNGNPAERNDIAMGPIINAAALFEVEKQVAVAEAGARVAFGKAVEG 360
 Db 401 FVNRLEFAMQAVQFNGNPAERNDIAMGPIINAAALFEVEKQVAVAEAGARVAFGKAVEG 360
 QY 461 KGYVPTPTLLDVRQKMSIMHEETFGPVLPVVAFTLEDAISMANDSDYGLTSSITDQ 420
 Db 461 KGYVPTPTLLDVRQKMSIMHEETFGPVLPVVAFTLEDAISMANDSDYGLTSSITDQ 420

Db 361 KGYVPTPTLLDVRQKMSIMHEETFGPVLPVVAFTLEDAISMANDSDYGLTSSITDQ 420
 QY 421 NVAMKAIKGLKFGFTYINRENFAMQGFHAGMKKSGTGGADCKHGLHFYLOTQVYVLOS 479
 Db 421 NVAMKAIKGLKFGFTYINRENFAMQGFHAGMKKSGTGGADCKHGLHFYLOTQVYVLOS 479
 RESULT 3
 Query Match 99.28; Score 2426; DB 2; Length 479;
 Best Local Similarity 99.48; Pred. No. 1,76-161;
 Matches 476; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 A:Species: *Escherichia coli*
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: D85737
 R:Perera, N.T.; Plunkett, H.; G. Pohlman, V.; Mau, R.; Garsner, J.P.; Rose, D.J.; May
 Miller, L.; Grothbeck, E.T.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoustis, K.; Apoda
 Nairo, A.; 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
 A:Reference number: A85480; MUID:21074945; PMID:11206551
 A:Accession: D85737
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1,479 <STO>
 A:Cross references: GB:AE005174; NID:q12512294; PDB:AA55636.1; GSPDB:GN00145; OM:FE
 A:Experimental source: strain 0157:H7, substrain EDL933
 C:Genetics:
 A:Gene: aldA
 C:Superfamily: aldohyde dehydrogenase (NAD+); aldohyde dehydrogenase homology

Query Match 99.28; Score 2426; DB 2; Length 479;
 Best Local Similarity 99.48; Pred. No. 1,76-161;
 Matches 476; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MSVPVQHPMTDGGFTWGRDAMIIVVNPATFAVISRIIPDQAFARATDAEAPQEW 60
 Db 1 MSVPVQHPMTDGGFTWGRDAMIIVVNPATFAVISRIIPDQAFARATDAEAPQEW 60
 QY 61 EALPAIERASWMLKISAGIREBASISALIVEGKIQOLAEVVAFTADYIDYAEAR 120
 Db 61 EALPAIERASWMLKISAGIREBASISALIVEGKIQOLAEVVAFTADYIDYAEAR 120
 QY 121 RYGEETIOSDRGENILLFKRALGYTTGILPNPFEPFLARKMAFALLTGNTIVIKSEF 180
 Db 121 RYGEETIOSDRGENILLFKRALGYTTGILPNPFEPFLARKMAFALLTGNTIVIKSEF 180
 QY 181 KYEETIIOSDRGENILLFKRALGYTTGILPNPFEPFLARKMAFALLTGNTIVIKSEF 180
 Db 181 KYEETIIOSDRGENILLFKRALGYTTGILPNPFEPFLARKMAFALLTGNTIVIKSEF 180
 QY 241 AAKNITKVCLEIGGKAPATVMDADDELAVKATVDSRVINSQVCCACARVYVVGITDQ 300
 Db 241 AAKNITKVCLEIGGKAPATVMDADDELAVKATVDSRVINSQVCCACARVYVVGITDQ 300
 QY 301 FVNRLEFAMQAVQFNGNPAERNDIAMGPIINAAALFEVEKQVAVAEAGARVAFGKAVEG 360
 Db 301 FVNRLEFAMQAVQFNGNPAERNDIAMGPIINAAALFEVEKQVAVAEAGARVAFGKAVEG 360
 QY 361 KGYVPTPTLLDVRQKMSIMHEETFGPVLPVVAFTLEDAISMANDSDYGLTSSITDQ 420
 Db 361 KGYVPTPTLLDVRQKMSIMHEETFGPVLPVVAFTLEDAISMANDSDYGLTSSITDQ 420
 QY 421 NVAMKAIKGLKFGFTYINRENFAMQGFHAGMKKSGTGGADCKHGLHFYLOTQVYVLOS 479
 Db 421 NVAMKAIKGLKFGFTYINRENFAMQGFHAGMKKSGTGGADCKHGLHFYLOTQVYVLOS 479

RESULT 4

Query Match 99.28; Score 2426; DB 2; Length 479;
 Best Local Similarity 99.48; Pred. No. 1,76-161;
 Matches 476; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 A:Species: *Neisseria meningitidis*
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
 C:Accession: A81023
 R:Peterson, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
 Hickey, P.K.; Hatt, D.H.; Salzberg, S.L.; White, O.; Foutschmann, P.D.; Dougherty, R.

RI, H.: Qiu, H.; Yamahewyan, J.; Gill, J.; Scariato, V.; Masignani, V.; Pizzo, M.
 Science 287, 1809-1815, 2000
 A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, G.M.; Moxon, E.P.; Rappelli, P.; Va
 A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MO58.
 A:Reference number: AB10007; MDL:20175755; PMID:10710307
 A:Accession: AB1023
 A:Molecule type: DNA
 A:Status: preliminary
 A:Residues: 1-480 <TEXT>
 A:Cross-References: GR AB002544, GR AE002068, NID: 37227214; PDB AAH42297; EMBL:3722724
 A:Experimental source: serogroup B, strain MO58
 C:Genetics:
 A:Gene: NMB1968
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 63.7%; Score 1557; DB 2; Length 480;
 Best Local Similarity 63.2%; Pred. No. 6,6e-101;
 Matches 301; Conservative 63; Mismatches 110; Indels 2; Gaps 1;

6 GHPMTTGGCTTWMKSLAMIVNPAFAVISPFLGDAEDAKKADAAERAPENELPA 65
 3 QIAMTINPFDENRGENPVLNPTFFAIAPFKKKAIVDPAAAPAAQDAPRFLPA 62
 66 IEFASMLKRIASGIREFASLSALIVEDECKIKQLAFVAPPAIADYVMAFAPPEYGE 125
 63 VEPGATYIKTAQGIPEPADELITIVAGGSKTKOLAVFVPMFADLYGAEWAPPEYGE 122
 126 IIGSPRENILTKRAIGVTTGILPNNPFLIARKMAALLTGNTIVIKPSEFTTNA 185
 123 IIGSPRENILTKRAIGVTTGILPNNPFLIARKMAALLTGNTIVIKPSEFTTNA 182
 186 IAFKIVDEIGLPGVFNVLVGRGETVGOELAGNPKAVMSKTSVAGKIMATAKNI 245
 183 HIFALIVAGLIPNVEHVGKALGELNLSAHPVDMSLTGVSAGQWMAASANI 242
 246 TKVLELGGKAPAVIMQADLELAVKALVUSPVINSQVNCAPVYVVGKIVDQVNRK 305
 243 TKVSELDGKAPAVIMQADLELAVKALVUSPVINSQVNCAPVYVVGKIVDQVNRK 302
 306 GFAAMAVQGNP--AEKNDIAMSPLINAALEPEYKVAAPVAVGAAVAGKAVBGKY 363
 303 TAMAGVYGNAPAEVAGALEDMGPIERBAKAAEVAAPVAVGAAVAGKAVBGKY 362
 364 YFPPLLDVROEKSIMHEETFGVLPVAFDTLEDAISMANSDVCLTSIYTNLNA 423
 363 FEEPLLDVROEKSIMHEETFGVLPVAFDTLEDAISMANSDVCLTSIYTNLNA 422
 424 MKAIKGLKFGELYINRENFEMAGHAGKSGIGAGDKHGLHGYQTQVYVLOS 479
 423 FYVTRIGFELYINRENFEMAGHAGKSGIGAGDKHGLHGYQTQVYVLET 478

RESULT 5

F81965
 Lactaldehyde dehydrogenase (EC 1.2.1.22) A NMA0480 [imported] *Neisseria meningitidis*
 C:Species: *Neisseria meningitidis*
 C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 05-Feb-2001
 C:Accession: F81965
 R:Parhill, J.; Achtman, M.; James, K.P.; Borthey, S.P.; Churcher, C.; Klemm, S.P.; Mout
 ; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
 Nature 404, 502-506, 2000
 A:Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* 22491.
 A:Reference number: AB17767; MDL:202256; PMID:10761305
 A:Accession: F81965
 A:Molecule type: DNA
 A:Status: preliminary
 A:Residues: 1-480 <PAR>
 A:Cross-References: GR A1112753, GR A1177959, NID: 37279120; PDB CAB87774; EMBL:3727922
 A:Experimental source: serogroup A, strain 22491
 C:Genetics:
 A:Gene: aldA; NMA0480
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 C:Keywords: oxidoreductase

Query Match 63.5%; Score 1552; DB 2; Length 480;
 Best Local Similarity 63.0%; Pred. No. 1.5e-100;
 Matches 300; Conservative 64; Mismatches 110; Indels 2; Gaps 1;

6 GHPVTLGELVFWKSLAMIVNPAFAVISPFLGDAEDAKKADAAERAPENELPA 65
 3 QIAMTINPFDENRGENPVLNPTFFAIAPFKKKAIVDPAAAPAAQDAPRFLPA 62
 66 IEFASMLKRIASGIREFASLSALIVEDECKIKQLAFVAPPAIADYVMAFAPPEYGE 125
 63 VEPGATYIKTAQGIPEPADELITIVAGGSKTKOLAVFVPMFADLYGAEWAPPEYGE 122
 126 IIGSPRENILTKRAIGVTTGILPNNPFLIARKMAALLTGNTIVIKPSEFTTNA 185
 123 IIGSPRENILTKRAIGVTTGILPNNPFLIARKMAALLTGNTIVIKPSEFTTNA 182
 186 IAFKIVDEIGLPGVFNVLVGRGETVGOELAGNPKAVMSKTSVAGKIMATAKNI 245
 183 HIFALIVAGLIPNVEHVGKALGELNLSAHPVDMSLTGVSAGQWMAASANI 242
 246 TKVLELGGKAPAVIMQADLELAVKALVUSPVINSQVNCAPVYVVGKIVDQVNRK 305
 243 TKVSELDGKAPAVIMQADLELAVKALVUSPVINSQVNCAPVYVVGKIVDQVNRK 302
 306 GFAAMAVQGNP--AEKNDIAMSPLINAALEPEYKVAAPVAVGAAVAGKAVBGKY 363
 303 TAMAGVYGNAPAEVAGALEDMGPIERBAKAAEVAAPVAVGAAVAGKAVBGKY 362
 364 YFPPLLDVROEKSIMHEETFGVLPVAFDTLEDAISMANSDVCLTSIYTNLNA 423
 363 FEEPLLDVROEKSIMHEETFGVLPVAFDTLEDAISMANSDVCLTSIYTNLNA 422
 424 MKAIKGLKFGELYINRENFEMAGHAGKSGIGAGDKHGLHGYQTQVYVLOS 479
 423 FYVTRIGFELYINRENFEMAGHAGKSGIGAGDKHGLHGYQTQVYVLET 478

RESULT 6

E81394
 probable lactaldehyde dehydrogenase (EC 1.2.1.22) truncated homolog C10490 [similarit
 C:Species: *Campylobacter jejuni*
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
 C:Accession: E81394
 R:Parhill, J.; Wren, R.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Klemm, S.P.; Ch
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVlier, A.; Whitehead, S.; Bar
 Nature 403, 665-668, 2000
 A:Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals
 A:Reference number: AB1250; MDL:2015312; PMID:10686204
 A:Accession: E81394
 A:Molecule type: DNA
 A:Status: preliminary
 A:Residues: 1-393 <PAR>
 A:Cross-References: GR A1130075, GR A1111168, NID:36967817; PDB CAB7128; EMBL:3696
 A:Experimental source: serotype O2, strain NCTC 1168
 C:Genetics:
 A:Gene: aldA; C10490
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 C:Keywords: oxidoreductase

Query Match 52.3%; Score 1270.5; DB 2; Length 393;
 Best Local Similarity 53.1%; Pred. No. 1.1e-81;

Matches 245; Conservative 52; Mismatches 90; Indels 1; Gaps 1;
 91 VEEGKIKQLAFVAPPAIADYVMAFAPPEYGEIIGSPRENILTKRAIGVTTGIL 150
 1 MEEGKIKQLAFVAPPAIADYVMAFAPPEYGEIIGSPRENILTKRAIGVTTGIL 60
 151 PNNPFLIARKMAALLTGNTIVIKPSEFTTNAIAFAIVDEIGLPGVFNVLVGRK 210
 61 PNNPFLIARKMAALLTGNTIVIKPSEFTTNAIAFAIVDEIGLPGVFNVLVGRK 120
 211 TVGELANPVAAPVAVGAAVAGKAVBGKYIMATAKNIIVTLEFSGKAPAVIMQADLELA 270

```

Db 121 VVGEELSSNNELGNVSLTGVPAATRTMFAAANKITIKVSLDGGKAPALVCRADITDLAV 180
OY 271 KALVDVSVINSQVCR AEEVYVQKGIYDQVNRKLCGMAQVQVCPARNDIAMPGLIN 340
Db 181 EAKASHIENNGVQVNCNFAVAVYTSVYDEFVQKFPVAKSKVSVNTI-KGFEDMGPLVN 239
OY 441 AALAEERVQKVARAVEGAVARVAFGKAVEGKYVYPPTLLVROBMSIMHEETGSPVLP 390
Db 240 GAVVDNMLAMIGATATKAGALVVGQKLTDTSTGYFPASVLTNKKHDEIMQKEITAPILP 299
OY 491 VVAHTLEDAISMANSDVCLTSSITVTUNLVAMKAIKIGKEGETYINRENEAMOGPIIA 450
Db 400 IAKEDTIDEVIMANKKEEYGLTSSITVTUNIDIMRARSREIKGETYINRENEAMOGPIIA 359
OY 451 GMRSGCTZANCKKHCICGYQCVVYIQ 478
Db 460 GFRKSHIGAMGKKGLEEVYLAHVYIQ 487

```

RESULT 7

```

Db 4613 succinate-semialdehyde dehydrogenase PA0265 [imported] Pseudomonas aeruginosa (strain
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: DB4613
C:Stover, C.K.; Pham, X.Q.; Ewlin, A.L.; Mazonuehi, S.D.; Warren, F.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Forder, K.R.; Kas, A.; Iatig, K.; Lim,
C.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: DB4613
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-483 <STAT>
A:Cross-references: GR:AE004464; GR:AE004091; NID:q9946099, P1DN AAF03664 1, GSPFR:GM001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: qabD
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

```

```

Query Match 46.08; Score 879.5; Db 2: Length 483;
Best Local Similarity 49.38; Pred. No. 1,2e-53;
Matches 184; Conservative 96; Mismatches 187; Indels 1; Gaps 1;
OY 10 YLDCQVYTWKQDAMIVVNPATFAVTSRIPDQAEIARAKAIJAAEKAQPEWALPAIEKA 69
Db 14 YVKKAVWDIUNQOTIKVNNPATGELIGSVKMKAAETBRATEADUKALPAMRALTAKEHA 73
OY 70 SWLKRTSALIRPPASFSALIVPEQGRIGQLAEVAVFALVITDMWAKHRRYEGFTTQS 129
Db 74 NKLRFWEDMTENQDILAKMLTDEQKPLAEAKGELAAVASPEEMEGEAKRIYDITIPG 133
OY 140 DREGENILFEKRALGVTTGILPMNPEPLIARKMAVALTGNTIVIKSEFTNNNAIFA 189
Db 144 HQRQKRLIVIKPIGTAAITTMNPFSAITTRKAGIALAAGCTIMWLKASQTFPSALALA 193
OY 190 KIVDEIGLRGVNLVLCGFTVGOELACNPVAVVSMGTSASAGEKIMATAAKNITKVC 249
Db 194 ELAEFRATIPKGVSVVTSAGCEVGETLSNPITPKLTFTGSTEIRQIMAFADQIKKVS 253
OY 250 LELGKAPIMVMDALIELAVKAVISKVINSGVQVNCNFAVAVYTSVYDQVNRKLCGMA 409
Db 254 LELAGNAPPEVVDADILMAVEGALISKYRNKGQVTCVANKLYVQDGVDAEVDKILAAV 313
OY 410 QAVQGNFAERNIDAMPILINAAALERVQKVARAVEGKAVEGKYVYPPTL 369
Db 414 AKININQNGLEAG-VITGPIIDAKAAKAVEFHIALDPAVSKAKVSGKTHALGTFEFTI 372
OY 470 LILVQKPKSLMHEETFGVLPVVAFDLLEDAISMANDSYGLTSSITVTUNLVAMKAIKG 429
Db 473 LIVVPPNNAIVSKDEPGLAPVPERKDEAFVITAMSNDEFTGLASVFARDILAPVFAVQ 432

```

```

OY 430 LKRGFTYINPNEFAMQIFPAKPKSKJGSAIKRHLRYIGTVYVIL 477
Db 433 LEYQVGINGLISNEVAPGEGIKASGTKEGSKYLEYDLIKYCL 480

```

RESULT 8

```

Db 65045 succinate-semialdehyde dehydrogenase [NAD(P)] (EC 1.2.1.16) - Escherichia coli (strain
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 03-Jun-2002
C:Accession: F65045
C:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
-A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:9742617; PMID:9278503
A:Accession: F65045
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-482 <BLAT>
A:Cross-references: GR:AE000361; GR:U00096; NID:q178911; P1DN AAF75708 1; P1D:q17890
A:Experimental source: strain K-12, Substrain MG1655
C:Genetics:
A:Gene: qabD
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
C:Keywords: oxidoreductase

```

```

Query Match 35.78; Score 873.5; Db 2: Length 482;
Best Local Similarity 39.58; Pred. No. 3e-53;
Matches 182; Conservative 91; Mismatches 187; Indels 1; Gaps 1;

```

```

OY 11 IDGQVYTWKQDAMIVVNPATFAVTSRIPDQAEIARAKAIJAAEKAQPEWALPAIEKAS 70
Db 15 INGMWLDANNGEALIDVTNPANGDKIGSVKMGKADETBRALIDANRALPAMRALTAKEKAT 74
OY 71 WLKRSALIRPPASFSALIVEEYKLYLAEEVAVFALVITDMWAKHRRYEGFTTQS 130
Db 75 TLKRFWEDMTENQDILAKMLTDEQKPLAEAKGELAAVASPEEMEGEAKRIYDITIPG 134
OY 131 RPEENILFEKRALGVTTGILPMNPEPLIARKMAVALTGNTIVIKSEFTNNNAIFA 190
Db 135 QAKRKRLIVIKPIGTAAITTMNPFSAITTRKAGIALAAGCTIMWLKASQTFPSALALA 194
OY 191 LVEIGLRGVNLVLCGFTVGOELACNPVAVVSMGTSASAGEKIMATAAKNITKVC 250
Db 195 IATIRAGVPAGVFNVTQSGAAGQNEITSNPLVKLTFTGSTEIRQIMAFADQIKKVS 254
OY 251 ELGKAPATYMDADIELAVKAVISKVINSGVQVNCNFAVAVYTSVYDQVNRKLCGMA 410
Db 255 ELGQNAPEVVDADILMAVEGALISKYRNKGQVTCVANKLYVQDGVDAEVDKILAAV 414
OY 411 AVQGNFAERNIDAMPILINAAALERVQKVARAVEGKAVEGKYVYPPTL 470
Db 415 KLHIGQGLD-NVITGPIIDAKAAKAVEFHIALDPAVSKAKVSGKTHALGTFEFTI 373
OY 471 LVRQKPKSLMHEETFGVLPVVAFDLLEDAISMANDSYGLTSSITVTUNLVAMKAIKG 430
Db 474 VDVPAKAKSKETFGPLAPLPERFKDEADVIQAANDTFGLAAVFYARDLSRVFVQAL 433
OY 431 KFGETYINRENEAMOGPIIAKPKSKJGSAIKRHLRYIGTVYVIL 471
Db 434 EYQVGINGLISNEVAPGEGIKASGTKEGSKYLEYDLIKYCL 474

```

RESULT 9

```

Db 891069 succinate-semialdehyde dehydrogenase [imported] - Escherichia coli (strain 0157:H7, S
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: B91069
C:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.
gasawara, N.; Yasunaga, T.; Kohara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

```

DNA Ref. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gene
 A:Ref:Gen. 8, 11-22, 2001. M010:21156231. PMID:11208746
 A:Accession: F01969
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1,483,144
 A:Cross-references: GR:BA000007; FIDN:HA16445.1; PID:01336293; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain RIMD 0509952
 C:Genetics: EC83522
 C:Superfamily: aldehyde dehydrogenase (NAD+), aldehyde dehydrogenase homology

Query Match 35.6%; Score 870.5; DB 2; Length 482;
 Best Local Similarity 39.3%; Pred. No. 4 9a-53;
 Matches 181; Conservative 92; Mismatches 187; Indels 1; Gaps 1;

OY 11 IDGQVYWRGDAMIDVNPATEAVISRIIPDQAFAPKATIDAEAPQEPALPAIRAS 70
 Db 15 INCEMLDANNGEVIDVTPNANDKIGSVKMGADETRAIDAANRALPVMALIAKERAN 74
 OY 71 WLKRTSAGIRERASEISALIVEEGKIQQLAEVAFATADYDYMAEMARREGELIOS 130
 Db 75 ILRNENLMMEHQDLARLMTLEGGKPLAEKAGETSYAASFTEWFAEGRKTYGDTIPGH 134
 OY 131 RGENILLFKRALGYTTGILPWNPPFLIARKMAPALLTGNTIYKSEFTNNALAFK 190
 Db 135 QADKRLIVKOPITVTAITTPNFPAMITRKAGALAGSTWIKRPSQSPSALALAE 194
 OY 191 IYDEIGLPGVENVLNGEIVGDELACNPVAVSMGTSVASEKIMATAKNITVCL 250
 Db 195 LAIRAGIPAGVENVTGSAVGNETSPLVRLSPSTETISPLMLHQCARKDICKVSL 254
 OY 251 ELGGRAPALVMDALALELAVKAIYDSKYNISQVNCNAPVYVKGITDYOVNLSGAMQ 310
 Db 255 ELGGNAPFTVFDALDAVEBQALASKRPNAGTVCANPLYVQGVYDRPAEKLDQAVS 314
 OY 311 AVQGNPAERNDIAMSPILMAALEEVEQKVAFAVEEAGVAFKAVFGKGYYPPTLL 370
 Db 315 KLHIGGLDKG-VTIGPLIDEKAVAKVEHIDALEKARAVCGKAKERGNFFOTIL 373
 OY 371 LDVROEMSLMEETGPPVLPVVAFTLEDALISMANDSYGLTSSTYTONLVAMKAIKGL 430
 Db 374 VVPANAKVSKETGGLAPLRFKDEADVAQNDIEFGAAVFAFTLSRVRVGEAL 433
 OY 431 KPGFTYINPENPEAMQSPHAGWPKSGIGAGDCKHGLHYIQ 471
 Db 434 EYGIYVINTGIIISNEVAPFGGIKASGLRGSKIGIEDYLE 474

RESULT 10
 D85913
 succinate-semialdehyde dehydrogenase [imported] - *Escherichia coli* (strain O157:H7, subs
 C:Species: *Escherichia coli*
 C:Date: 11-Feb-2001 #sequence_revision 16 Feb 2001 #text_change 27-Nov-2001
 C:Accession: D85913
 R:Perma. N.T. Plunkett III, G. Paulsard, V. Mau, R. Gibson, J.P. Rose, D.J. Mayhew
 Hilder, L.J. Grobeck, E.J. Davis, N.W. Lim, A.J. Dimalanta, E.J. Polomousis, K.J. Apodaca
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
 A:Reference number: AB04480; M010:21074935; PMID:11208551
 A:Accession: D85913
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1,483,144
 A:Cross-references: GR:AE005174; NIT:012517054; FIDN:AA157768.1; GSPDB:GN00145; OMIM:244
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 C:Gene: 9abD
 C:Superfamily: aldehyde dehydrogenase (NAD+), aldehyde dehydrogenase homology

Query Match 35.6%; Score 870.5; DB 2; Length 482;
 Best Local Similarity 39.3%; Pred. No. 4 9a-53;

Matches 181; Conservative 92; Mismatches 187; Indels 1; Gaps 1;

OY 11 IDGQVYWRGDAMIDVNPATEAVISRIIPDQAFAPKATIDAEAPQEPALPAIRAS 70
 Db 15 INCEMLDANNGEVIDVTPNANDKIGSVKMGADETRAIDAANRALPVMALIAKERAN 74
 OY 71 WLKRTSAGIRERASEISALIVEEGKIQQLAEVAFATADYDYMAEMARREGELIOS 130
 Db 75 ILRNENLMMEHQDLARLMTLEGGKPLAEKAGETSYAASFTEWFAEGRKTYGDTIPGH 134
 OY 131 RGENILLFKRALGYTTGILPWNPPFLIARKMAPALLTGNTIYKSEFTNNALAFK 190
 Db 135 QADKRLIVKOPITVTAITTPNFPAMITRKAGALAGSTWIKRPSQSPSALALAE 194
 OY 191 IYDEIGLPGVENVLNGEIVGDELACNPVAVSMGTSVASEKIMATAKNITVCL 250
 Db 195 LAIRAGIPAGVENVTGSAVGNETSPLVRLSPSTETISPLMLHQCARKDICKVSL 254
 OY 251 ELGGRAPALVMDALALELAVKAIYDSKYNISQVNCNAPVYVKGITDYOVNLSGAMQ 310
 Db 255 ELGGNAPFTVFDALDAVEBQALASKRPNAGTVCANPLYVQGVYDRPAEKLDQAVS 314
 OY 311 AVQGNPAERNDIAMSPILMAALEEVEQKVAFAVEEAGVAFKAVFGKGYYPPTLL 370
 Db 315 KLHIGGLDKG-VTIGPLIDEKAVAKVEHIDALEKARAVCGKAKERGNFFOTIL 373
 OY 371 LDVROEMSLMEETGPPVLPVVAFTLEDALISMANDSYGLTSSTYTONLVAMKAIKGL 430
 Db 374 VVPANAKVSKETGGLAPLRFKDEADVAQNDIEFGAAVFAFTLSRVRVGEAL 433
 OY 431 KPGFTYINPENPEAMQSPHAGWPKSGIGAGDCKHGLHYIQ 471
 Db 434 EYGIYVINTGIIISNEVAPFGGIKASGLRGSKIGIEDYLE 474

RESULT 11
 AE0839
 succinate-semialdehyde dehydrogenase (NAD(P)) (EC 1.2.1.16) - *Salmonella enterica* sub
 C:Species: *Salmonella enterica* subs. *enterica* serovar Typh
 A:Note: this species has also been called *Salmonella typhi*
 C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 03-Jun-2002
 C:Accession: AE0839
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.P.; Pickard, D.; Wain, J.; Church
 H.; T. J. Connerton, P.; Cronin, A.; Davis, P.; Davies, P.M.; Dowd, L.; White, N.; Fair
 S.; Mout, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A:Title: The genome of *Salmonella enterica* serovar Typhimurium.
 A:Reference number: AB0502; PMID:11677608
 A:Accession: AE0839
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1,482,474
 A:Cross-references: GR:AL11342; FIDN:AA05000.1; FID:01336294; GSPDB:GN00176
 C:Genetics: STY2911
 C:Superfamily: aldehyde dehydrogenase (NAD+), aldehyde dehydrogenase homology
 C:Keywords: oxidoreductase

Query Match 35.1%; Score 858.5; DB 2; Length 482;
 Best Local Similarity 39.2%; Pred. No. 4 4a-52;
 Matches 181; Conservative 91; Mismatches 189; Indels 1; Gaps 1;

OY 10 YHGGQVYWRGDAMIDVNPATEAVISRIIPDQAFAPKATIDAEAPQEPALPAIRAS 69
 Db 14 ILHGLVWIKAKAGSIVLVNPNANAKRLSNVPPMGAFPPADINANRALPVMALIAKERAN 73
 OY 70 WLKRTSAGIRERASEISALIVEEGKIQQLAEVAFATADYDYMAEMARREGELIOS 129
 Db 74 ILRNENLMMEHQDLARLMTLEGGKPLAEKAGETSYAASFTEWFAEGRKTYGDTIPGH 133
 OY 130 RGENILLFKRALGYTTGILPWNPPFLIARKMAPALLTGNTIYKSEFTNNALAFK 189

Db 134 HODTRRLVATKQFVIAATITPWNPTSAMITTKKAPALACGTMTAKPSQTPPSALMA 193
 Oy 190 KIVETELDRCVNLVLSFETVGGELACNPKVAMVSTGVSACRIMATAKNITVGC 249
 Db 194 ELAARACILAGVENVVGSADIGDELSTNPIVKRLSTFISTEIGKQIMEGCAKDKRVS 253
 Oy 250 LELAGKAPALVMDADDELAVKALVDSRVINSQVNCAPRYVOKGIDYOVNKLGEAM 309
 Db 254 LELAGNAPITVTDADILAKRAGCALASKERNACQTTVCANRILYVDQVDRFAEKLNQAV 313
 Oy 410 QAVQFQNPARKNIIAMQPIINAAALERVOKVAVRERAGAVAFSGKAVEGKGYPPTL 369
 Db 414 NKLAVRQKICQAVATGPELDEKAVAKVGEHIDALEKCARVITGGEAKILGNGFEPTI 372
 Oy 470 LILVROFMSIMHEFTGVPVLYVAFITLIDALISMANDSYGCTSSITYONLNVAKAKIG 429
 Db 473 LAIVPDKAKAKKEETPDLAFIRFESDEADVIRQANTTFGLAAVYADNHSVPVQCPA 432
 Oy 440 LKRETTYINRENEAMOGFHAMGPKSSTGATACRKHGLGYLQ 471
 Db 443 LEVGLVGTINTGILISNEVAPPGCISKAGLGRSKGYTIDYLE 474

RESULT 12
 Db4064
 sucinate-semialdehyde dehydrogenase gabd [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C:Accession: DB4064
 R:Tokami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
 Nucleic Acids Res 20: 4417-4431, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
 A:Reference number: AB4650; MIMD:20512582; PMID:11058132
 A:Accession: DB4064
 A:Status: preliminary
 A:Molecule type: DNA
 A:Keywords: 1-475, str.
 A:Cross references: GB:AF00151b; GB:BA000004; NID:q10175/94; PIDN:BAH0705-1; GSPDB:GN00
 A:Experimental source: strain C-125
 C:Genes:
 C:Gene: gabd
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 44.68; Score 845.5; Db 2; Length 475;
 Best local similarity 39.18; Prod. No. 2.7e-51;
 Matches 184; Conservative 84; Mismatches 195; Indels 7; Gaps 4;

Oy 9 MTITGQFVIMGCA--WIDVNPATEAVISRLPDGADAPKATIAERAPQEPWALPAI 66
 Db 6 LITDG---TWTGILLDTPVKNPATGEVYGVMPNGKAEAAALTAQARAFIKRRTAA 62
 Oy 67 ERASWLRKISAGIRERASEISALIVERGKIQGLAVVYAPTADYIDYMAFWAPPYGEI 126
 Db 63 ERANPTMTWIDLLQCKKEELAEIMTEMKRIPLAFAGLEIYASPIPMFAEGKRVYERT 122
 Oy 127 IQSDRGENTILFKRALVGTGILLPWNPEFLIARKMAPALLTGMTIYIKSEPTNNAI 186
 Db 123 VASASAKRKTHVQKQVGVAAITPWNFPAAMIVKMAALAAAGTEFGKAEILPLIPLTA 182
 Oy 187 AFAKIVDEIGLPRGVNLVLSRGTEVQGLACNPKVAMVSTGVSASCEKIMATAKNIT 246
 Db 183 KIVELDEEAFPRGVNVAWIRAKLIGFEMTSHEIVPKLETTGSAVSKLIMKSGASGML 242
 Oy 247 KVLLEAGKAPALVMDADDELAVKALVDSRVINSQVNCAPRYVOKGIDYOVNKLGE 306
 Db 243 NLSLEAGQAPMLICDADICRAVAVASAKRYNMGQTVGNRIYVDETTIVDEPKELT 302
 Oy 307 EAMQVQFQNPARKNIIAMQPIINAAALERVOKVAVRERAGAVAFSGKAVEGKGY 365
 Db 303 GKAVVQIKVNGILFEGVHTIGPLIEKKGYKKAVIYDQAVAKARAVIGGKQKQGNDSFY 361
 Oy 366 PPTLLADVQKMSIMHEFTGVPVLYVAFITLIDALISMANDSYGCTSSITYONLNVAK 425

Db 362 DPTLLDVHDEMLVMOELTFGVAPVIGTATVEDVIFRANCTRYGLAVFTENTARGL 421
 Oy 426 AIKQKRETYINRENEAMOGFHAMGPKSSTGATACRKHGLGYLQ 475
 Db 429 LSPALDPGICVGNMGCAFSLAQAPFGSKESGLRFRVQGGELFALETFKV 471

RESULT 13
 C95948
 Probable succinate-semialdehyde dehydrogenase [NAD(P)] (EC 1.2.1.16) [imported] - Sin
 C:Species: Sinorhizobium meliloti
 C:Date: 24 Aug 2001 #sequence_revision 24 Aug 2001 #text_change 03 Jun 2002
 C:Accession: C95948
 R:Finan, T.M.; Weidner, S.; Wong, K.; Bihmester, J.; Chain, P.; Vorholter, F.J.; Her
 Proc Natl Acad Sci U S A 98: 9890-9894, 2001
 A:Title: The complete sequence of the 1.685-kb psymb megaplasmid from the N2-fixing e
 A:Reference number: A95842; MIMD:21365508; PMID:11481431
 A:Accession: C95948
 A:Status: preliminary
 A:Molecule type: DNA
 A:Keywords: 1-491, KUR

A:Cross references: GB:AL521985; PIDN:CAQ49251.1; PID:q15140747; GSPDB:GN0167
 A:Experimental source: strain 1021, megaplasmid pSymb
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hub
 pel, A.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, P.H.; Kiss, E.; Komp, C.; Lelau
 hehault, P.; Vandenholte, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,
 A:Title: the composite genome of the legume symbiont Sinorhizobium meliloti.
 A:Reference number: A96039; MIMD:21368234; PMID:11474104
 A:Contents: annotation
 C:Genes:
 A:Gene: gabd2; SMD21185
 A:Genome: plasmid
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology
 C:Keywords: oxidoreductase

Query Match 34.28; Score 836; Db 2; Length 491;
 Best local similarity 39.38; Prod. No. 1.3e-50;
 Matches 176; Conservative 85; Mismatches 185; Indels 2; Gaps 2;

Oy 25 DYNPATVAVISITPRGQAEARAKATDAERAPQEPWALPAIFASWLRKISAGIRERAS 84
 Db 36 DVNPSTGELATLIPMGIDDAKTAIDAAALQPLMAPKPAKDRSILIRWHDILVHAD 95
 Oy 85 EISALIVEEGKIQGLAEVEVAFADYIDYMAFWAPPYGEIISQSDRGENTILFKRALG 144
 Db 96 DLVALITAEKGRPVGAKGEVLHMAASYEMRIEERKRVYGETFPAPANDRMILYIKQPVG 155
 Oy 145 VITGILPWNPEFLIARKMAPALLTGMTIYIKSEPTNNAIAPAKIVDEIGLPRGVNL 204
 Db 156 VVGTITPWNFPAAMVARKISPLAAGCTVLEKPAQGTPLVAGAMPVLAERKAGFEGLNL 215
 Oy 205 V-LGRGETYGOELAGCPKAVVAMVSTGVSASGKIMATAKNITVGCLELGKAPALYMD 263
 Db 216 LVASGAPVIGRELCPKVRKISFTGSTRVGLLRKQCSDOIKKVSLDLSGNAPFIYVD 275
 Oy 264 ADLEAVKAIVDSRVINSQVNCAPRYVOKGIDYOVNKLGEAMVAVQNPAPRNDI 323
 Db 276 ADIDEAVDCAVQAKFPNACQTCVSNRIYVQSAVHDVAFRFRVPLVIGD-GFAPDV 334
 Oy 324 AMQPIINAAALERVOKVAVRERAGAVAFSGKAVEGKGYPPTLADVQKMSIMHE 383
 Db 335 AIGPMIDAHADIKTANVADAQKAOVSGSGRIIGTGTFPFTVLTGISHMKIAOE 394
 Oy 384 TEGVPLVVAQFLLDELALISMANDSYGCTSSITYONLNVAKAKILGKRETYINRENE 443
 Db 395 TREPILAPITRFTARQVAVANDTITGLAAVYADNHSVPVQCPAINTGRMS 454
 Oy 444 AMQFHAMGPKSSTGATACRKHGLGYLQ 471
 Db 455 SFMAPGSGIKGSGISGPRSSRHGLFDYLE 482

Result 14

C83774

Succinate-semialdehyde dehydrogenase BH0995 [Imported] - *Bacillus halodurans* (strain C-1)

C:Species: *Bacillus halodurans*

C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: C83774

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, P.; Masui, N.; Fujii, F.; Hirai

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and

A:Reference number: AB3650; MIMD:20512582; PMID:11058132

A:Accession: C83774

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1468 <STO>

A:Cross-References: GB:AF001510; GB:BA000004; MID:31017440; PIRN:BA004714.1; GSPDB:GN0148

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH0995

C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 34.1%; Score 835; DB 2; Length 468;

Best Local Similarity 36.4%; Pred. No. 1.4e-50;

Matches 170; Conservative 105; Mismatches 186; Indels 6; Gaps 3;

9 MYIDGFTWGDAMIDVNPATPAVLSRIPDGAEDARKAIDAEKAPFEMALPAER 68

3 LYINEMW--RSGKTLDVNPATGEVIDIVFPAKKAELAVSAVAEAFQWSSQVASK 60

69 ASWLKISAGIRERASEISALIVEGGKIQULAEVEVAFADYIDVMAEWARRYEGELIQ 128

61 SPYLMRWQIINDQODEIGELIMTEKQKPLREAFGEVQVANSFQWAEAKRIYGDTP 120

129 SDRESENTLLKRALCVTTGILPNNPPLIARKMAPALLTONTIYIKSEPTNNALAF 188

121 ASAIKRLIKOPKOPVIAATTPNNPFAAMITRKVAALAGCAIYKPAEQPTLAIKL 180

189 AKIVDEIGPRGVNLYVIGRGETVGOELGNPKVAVMSGVSAGEKIMATAKNTKY 248

181 AOLAEAGIPAGVNLVITGNMODIGEALEDSRKRTITFTGSTEYGLMLRGAQIVYKI 240

249 CLEIGKAPAIYMDADLELAKAIYDSRVAINSGQVCAERYVYKGIYDQFVNRLEGA 308

241 STEIGGAPETIIMDANIEEAVDQVIAKSEFNMAQTVVANKIYVAFTAEATKFAAK 300

309 MQAVQFGNPAERKNIDIAMGPIINAAALRKVRQKVAARAEKARAFGKAAEGKGYPT 368

301 VNEIKVNGLEFEG-VTIGPLIDKAAVEKFAHITHDAIKKGGQYTVGSP--WTNHEFEPT 356

369 LLLDVROEMSIHMETEGPVLVPAFPLEDAISMANDSYGLTSSITYGNTLVAMKAIK 428

357 IITGATDEMIGMNETEGPLAPVATFDTBEVIERANHTPYGLAAYFTENIGRAIRLSE 416

429 GLKSGETITNPNFEAMQGFHAKWPKSGTGGADSKHGLGYLOTQV 475

417 KLEVGIVGVNDMPSSVAQAPFGKMGKESGLGREGKGYGIEEYLEKVV 463

Result 15

B87638

Succinate-semialdehyde dehydrogenase [Imported] - *Caulobacter crescentus*

C:Species: *Caulobacter crescentus*

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001

C:Accession: B87638

R:Nierman, W.C.; Pajk-Lyuu, T.V.; Paulsen, I.T.; N-lisou, K.E.; Eisen, J.; Heidelberg, J.

B.; Lamb, M.T.; DeRuy, R.T.; Dodson, F.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, J.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of *Caulobacter crescentus*.

A:Reference number: AB7249; MIMD:21173698; PMID:11259647

A:Accession: B87638

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1482 <STO>

A:Cross-References: GB:AE005673; MID:313424806; PIRN:AAK25102.1; GSPDB:GN0148

C:Genetics:

A:Gene: CC3140

C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 33.6%; Score 822; DB 2; Length 482;

Best Local Similarity 39.4%; Pred. No. 1.2e-49;

Matches 185; Conservative 82; Mismatches 199; Indels 4; Gaps 3;

5 VHHKATLVGDFVTWKGSLAMLVNPATPAVLSRIPDGAEDARKAIDAEKAPFEMALPAER 64

7 VETALLIGQWV--RSGKTLDPNPATGEVIDIVFPAKKAELAVSAVAEAFQWSSQVASK 64

65 ALEASWLPKISASIPKASHISALIVEGGKIQULAEVEVAFADYIDVMAEWARRYEGELIQ 124

65 AKPRVATILPWSSTIILAHADQAPLMTIPQKPKIAFAKKEFVVYGAASFILMPAEKAPVAV 124

125 FIQSDPRESENTLLKRALCVTTGILPNNPPLIARKMAPALLTONTIYIKSEPTNNALAF 184

125 HILPTPMGKRLASIKQVGVGCAALAPNPEPLAMIIKKVSPALAAKCTIVVAKPAALPLS 184

185 AIAFAKIVDEIGLPRGVNLYVIGRGETVGOELGNPKVAVMSGVSAGEKIMATAKNTKY 243

185 ALATAPLATKACVAGVNLVITGNMODIGEALEDSRKRTITFTGSTEYGLMLRGAQIVYKI 244

244 NFINVTEIGLPRGVNLYVIGRGETVGOELGNPKVAVMSGVSAGEKIMATAKNTKY 303

245 TMKRLSTELGNAPEFIVEEDADLEAAVDGAIASKYRNAGQVCANRLIVGSGIHDAFAA 304

304 PLTEAMAVFGNPAERKNIDIAMGPIINAAALRKVRQKVAARAEKARAFGKAAEGKGYPT 363

305 RLAEKVAALKVG--POTGEGVQIGPLINERKALIKVGLVSGAVAGAVLIGDVGGLGCH 363

364 YYPPTLLDVROEMSIHMETEGPVLVPAFPLEDAISMANDSYGLTSSITYGNTLVAMKAIK 423

364 FYQPTVLVAGATPEKRIFOEELIFGVADIVFTEAEAEVELANATPGLAAYFEYSRDVGC 423

424 MKATKSKIKFSTYTNPNFEAMQGFHAKWPKSGTGGADSKHGLGYLOTQV 473

424 WKVAEQIEAGVNGINEGLISTEVAPFGGVKESGLGREGKGYGIEEYLEKVV 473

Search completed: June 24, 2004, 10:18:31
Job time: 19.3303 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw mode

Run on: June 24, 2003, 10:02:55; Search time: 6.4773 seconds

(without alignments)
2059,300 Million cell updates/sec

Title: us-09-830-751-6

Perfect score: 2446

ADKAGHGLTQTVYLOS 479

Sequence: 1 MSVPVQHPHMDGFTVTRG...

Scoring table: BLOSUM62

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	ID	Description
1	2427	99.2	478	1	ALDA_ECOLI	P25553 Escherichia
2	878.5	35.7	482	1	CABD_ECOLI	P25556 Escherichia
3	810	33.1	491	1	GABD_RHSN	P55653 Rhizobium s
4	804.5	32.9	490	1	PHAR_RACSN	P71016 bacillus su
5	783.5	32.0	474	1	YDCW_ECOLI	P77674 escherichia
6	777	31.8	519	1	PHAC_YEAST	P46367 saccharomyc
7	769	31.4	500	1	DHAL_AGABI	074187 agacharomyc
8	767	31.4	488	1	SSDH_RAT	P51650 rattus norv
9	759.5	31.1	535	1	DHAF_YEAST	P40047 saccharomyc
10	750	30.7	519	1	DHAF_YEAST	P40047 saccharomyc
11	728	29.8	498	1	DHAF_YEAST	P40047 saccharomyc
12	726	29.7	497	1	DHAL_LETTA	P25417 Leishmania
13	719.5	29.4	492	1	DHAF_YEAST	P47895 homo sapien
14	719	29.4	499	1	DHAF_YEAST	P47895 homo sapien
15	718.5	29.4	497	1	DHAF_YEAST	P47895 homo sapien
16	713	29.1	492	1	FTCH_RAT	P17202 sphincter m
17	710	29.0	500	1	DHAF_YEAST	P28037 rattus norv
18	708.5	29.0	487	1	DHAF_YEAST	P54232 rhizobium m
19	708.5	29.0	497	1	UCAS_YEAST	P38067 saccharomyc
20	708	28.9	489	1	DHAL_ECOLI	P17445 escherichia
21	708	28.9	497	1	DHAL_ECOLI	P17445 escherichia
22	702.5	28.7	493	1	DHAF_HUMAN	P41761 aspergillus
23	699.5	28.6	503	1	DHAF_HUMAN	P41761 aspergillus
24	699	28.6	502	1	FTDH_HUMAN	P56533 gadus calla
25	698	28.6	502	1	DHAF_HUMAN	P56533 gadus calla
26	696	28.4	500	1	DHAF_HUMAN	P42757 atirplex bo
27	694	28.4	506	1	DHAL_RAT	P59808 schistosom
28	690.5	28.2	497	1	DHAL_ECOLI	P51647 rattus norv
29	687.5	28.1	500	1	DHAL_ECOLI	P51647 rattus norv
30	687.5	28.1	519	1	DHAL_ECOLI	P51647 rattus norv
31	686.5	28.1	519	1	DHAL_ECOLI	P51647 rattus norv
32	685.5	28.0	488	1	DHAL_ECOLI	P51647 rattus norv
33	684.5	28.0	500	1	DHAL_ECOLI	P51647 rattus norv

34	684	28.0	500	1	DHAL_MOUSE	P24549 mus musculus
35	683.5	27.9	500	1	DHAL_SHEEP	P51977 ovis aries
36	681	27.8	496	1	DHAL_GLAHE	P43108 ciadospoiti
37	680.5	27.8	500	1	DHAL_HUMAN	P00352 homo sapien
38	678.5	27.7	500	1	DHAL_BOVIN	P48644 bos taurus
39	678	27.7	500	1	DHAL_BOVIN	P48644 bos taurus
40	677.5	27.7	517	1	DHAL_HUMAN	P05091 homo sapien
41	676.5	27.6	500	1	DHAL_HUMAN	P12762 equus caball
42	675	27.6	501	1	DHAL_HUMAN	P04795 arathidopsis
43	675	27.6	501	1	DHAL_HUMAN	P04795 arathidopsis
44	672	27.5	501	1	DHAL_ELEPH	P04839 elephantiu
45	671.5	27.5	499	1	FEAR_ECOLI	P06568 escherichia

ALIGNMENTS

RESULT 1
ALDA_ECOLI STANAPP: PRT: 478 AA
ID ALDA_ECOLI
AC P25553:
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aldehyde dehydrogenase A (EC 1.2.1.22) (Lytic aldehyde 3-hydrogenase).
GN ALDA OR ALD OR B415.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
RC STRAIN-K12:
RX MEDLINE=92011371; PubMed=1917845;
RA Hidalgo E., Chen Y.-M., Lin E.C.C., Aguilar J.;
RT "Molecular cloning and DNA sequencing of the Escherichia coli K-12
PL aId gene encoding aldehyde dehydrogenase.";
PL J. Bacteriol 173:6118-6123(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. II, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Verde J., Glasner J.P., Rode C.K., Mayhew G.F.,
RA Gregor T., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose P.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K 12.";
PL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12:
RX MEDLINE=97251357; PubMed=9037033;
RA Alta H., Kasai H., Fujita K., Hayashi K., Inada T., Isono K.,
RA Titch T., Basal T., Kashimura K., Kimura S., Kitakawa M.,
RA Motomura K., Nakano F., Miki T., Mitsuuchi K., Miki H., Miki T.,
RA Motomura K., Nakano F., Miki T., Mitsuuchi K., Miki H., Miki T.,
RA Oshima T., Saito J., Saito G., Seki Y., Sivasubramanian S.,
RA Tsujimura H., Takeda J., Takeda K., Takeda Y., Wada C.,
RA Yamamoto Y., Horikuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K 12 genome
PL corresponding to the 28.0-49.1 min region on the linkage map.";
PL DNA Res 3:363-377(1996).
RN [4]
RP SEQUENCE OF 1-11.
RC STRAIN K12 / EMCC;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.T., Robinson K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
PL in the genome of Escherichia coli K 12.";
PL Electrochimica Acta 44:1259-1313(1997).
CC - FUNCTION: ACTS ON LACTALDEHYDROGENASE AS WITH AS OTHER ALDEHYDES.
CC - CATALYTIC ACTIVITY (S)-lactaldehyde + NAD(+) -> H(2)O + (S)-
CC lactate + NADH.

```

CC -1 SUBUNIT: HOMOTETRAMER.
CC -1 INDUCTION: BY GROWTH ON FUCOSE, RHAMNOSE, ARABINOSE AND AMINO
CC ACIDS SUCH AS GLUTAMATE.
CC -1 SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M64541; AAC23427.1;
CC EMBL: AF000249; AAC74497.1;
CC EMBL: D90780; BAA15032.1;
CC EMBL: D90781; BAA15037.1;
CC PIR: A8165; A8165.
CC HSSP: P06534; 1A48.
CC SWISS_2DPAGE: P25553; COL1.
CC EC00900: EC10045; ALD4.
CC InterPro: IP002086; Aldehyde_dehydr.
CC Pfam: PF00171; aldehyd_1.
CC PROSITE: PS00687; ALDEHYDE_DEHYDR_CYS; 1.
CC PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
CC OXIDOREDUCTASE: NADP: complete proteome.
CC INT_MET 0
CC NP_BIND 206 212 NAD (ADP PART) (BY SIMILARITY).
CC ACT_SITE 250 250 BY SIMILARITY.
CC ACT_SITE 284 284 BY SIMILARITY.
CC SEQUENCE 478 AA; 52141 MW; 0H5DF4B29B874A CRC64;

Query Match 99.4% Score 2427; BH 1; Length 478;
Host: Local Similarity 99.6% Pval: 1.9e-155;
Matches 476; conservative 0; Mismatches 2; Indels 0; Gaps 0;

CC 1 SVYVQHMVYIDGQFVTRKQAMIDVNPATEVVISIKPIGQEDARKAIDAERKOPME 61
CC 2 SVYVQHMVYIDGQFVTRKQAMIDVNPATEVVISIKPIGQEDARKAIDAERKOPME 61
CC 1 SVYVQHMVYIDGQFVTRKQAMIDVNPATEVVISIKPIGQEDARKAIDAERKOPME 60
CC 62 ALPAIEKASWIKRISAGIEPRKASEISALIVEEGKIQOLAEVEVAFTADYIDYAEWARR 121
CC 61 ALPAIEKASWIKRISAGIEPRKASEISALIVEEGKIQOLAEVEVAFTADYIDYAEWARR 120
CC 122 YEGEIIQSDRPGENILFEKRALGVTTGILPMNPPEFLIARKKAPALLTGNTIVIKPSEFT 181
CC 121 YEGEIIQSDRPGENILFEKRALGVTTGILPMNPPEFLIARKKAPALLTGNTIVIKPSEFT 180
CC 1B2 TNNAIAFAKIVDEIGLPGVFNLYIGRGETVGOELAGNPKVAMVMTGVSNGEKIMATA 241
CC 1B1 PNNAIAFAKIVDEIGLPGVFNLYIGRGETVGOELAGNPKVAMVMTGVSNGEKIMATA 240
CC 242 AKKITVCTLEIGKKAIVKAGLGLFLAVKAVIVSKVINSGVNCMAKVVYVQKITYGF 301
CC 241 AKKITVCTLEIGKKAIVKAGLGLFLAVKAVIVSKVINSGVNCMAKVVYVQKITYGF 300
CC 402 VNRIEFAWQVQVGFNFAERNDIAMGPIINAAALEVEVEQVAVAEVEGARVAFGKAVGSK 361
CC 401 VNRIEFAWQVQVGFNFAERNDIAMGPIINAAALEVEVEQVAVAEVEGARVAFGKAVGSK 360
CC 462 GYVYPTLLDVQWQENSIMHEETFGVLTVAVFTLEAATISMANDSDYGLTSSITYQNLN 421
CC 461 GYVYPTLLDVQWQENSIMHEETFGVLTVAVFTLEAATISMANDSDYGLTSSITYQNLN 420
CC 422 VAMKAIEKIKPGCEVYNRENTFAMQSFIAAGMKKSGICAGCKKHGHLYLTGVVYVIGS 479
CC 421 VAMKAIEKIKPGCEVYNRENTFAMQSFIAAGMKKSGICAGCKKHGHLYLTGVVYVIGS 478

```

```

DT 01-MAY-1992 (Ref. 22, Created)
DT 01-MAY-1992 (Ref. 22, Last sequence update)
DT 15-JUN-2002 (Ref. 41, Last annotation update)
DT Succinate-semialdehyde dehydrogenase [NADP+] (EC 1.2.1.16) (SSDH).
DT GAND OR h2661.
DT Escherichia coli.
DT Escherichia coli.
DT Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
DT Escherichia.
DT NCBI_TaxId=562;
DT [1]
DT SEQUENCE FROM N.A.
DT STRAIN-K12 / JM103;
DT MEDLINE=94127927; PubMed=8297211;
DT Niegemann E., Schulz A., Bartsch K.;
DT "Molecular organization of the Escherichia coli gab cluster:
DT nucleotide sequence of the structural genes gabH and gabP and
DT expression of the GABA permease gene."
DT Arch Microbiol 160:454-460(1993).
DT [2]
DT SEQUENCE FROM N.A.
DT STRAIN-K12 / MG1655;
DT MEDLINE=97426617; PubMed=9278503;
DT Blatter F.P., Plunkett G., III, Bloch C.A., Petra N.T., Hurland V.,
DT Riley M., Collado-Vides J., Glasner J.D., Poole C.K., Mayhew G.F.,
DT Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
DT Mau B., Shao Y.;
DT "The complete genome sequence of Escherichia coli K-12."
DT Science 277:1233-1238(1997).
DT [3]
DT SEQUENCE FROM N.A.
DT STRAIN-K12;
DT MEDLINE=97349980; PubMed=9205837;
DT Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
DT Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
DT Mizubuchi K., Mori H., Nakeda S., Nakamura Y., Nishimoto H.,
DT Oshima H., Oshima S., Saito N., Sempel G., Satoh Y., Sivasubraman S.,
DT Tagami H., Takahashi H., Takeeda J., Takemoto K., Uehara K., Wada C.,
DT Yamagata S., Horiiuchi T.;
DT "Construction of a contiguous 874-kb sequence of the Escherichia coli
DT K-12 genome corresponding to 50,000,000 bp on the linkage map and
DT analysis of its sequence features."
DT DNA Res. 4:91-113(1997).
DT CC -1 CATALYTIC ACTIVITY: Succinate semialdehyde + NAD(P)(+) + H(2)O =
DT succinate + NAD(P)H.
DT CC -1 PATHWAY: 4-aminobutyrate (GABA) degradation.
DT CC -1 SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
DT -1 CAUTION: Ref.3 sequence differs from that shown due to frameshifts
DT in positions 51 and 190.
DT
DT -----
DT THIS SWISS-PROT entry is copyright. It is produced through a collaboration
DT between the Swiss Institute of Bioinformatics and the EMBL outstation -
DT the European Bioinformatics Institute. There are no restrictions on its
DT use by non-profit institutions as long as its content is in no way
DT modified and this statement is not removed. Usage by and for commercial
DT entities requires a license agreement (See http://www.isb-sib.ch/announce/
DT or send an email to license@isb-sib.ch).
DT
DT EMBL: M88334; AAC36831.1;
DT EMBL: AF000051; AAC75708.1;
DT EMBL: AF00890; BAA16522.1; ALT_FRAME.
DT EMBL: D90890; BAA16523.1; ALT_FRAME.
DT EMBL: D90890; BAA16524.1; ALT_FRAME.
DT HSSP: P51977; 1BXS.
DT Ecogeno: EC01329; gabD.
DT InterPro: IP002086; Aldehyde_dehydr.
DT Pfam: PF00171; aldehyd_1.
DT PROSITE: PS00687; ALDEHYDE_DEHYDR_CYS; 1.
DT PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
DT OXIDOREDUCTASE: NADP: Complete proteome.
DT NP_BIND 233 238 NAD(P) (ADP PART) (BY SIMILARITY).
DT ACT_SITE 255 255 BY SIMILARITY.
DT ACT_SITE 289 289 BY SIMILARITY.
DT SEQUENCE 482 AA; 51720 MW; 0Y153F8F7410B0C4 CRC64;

```

```

Query Match: 35.7% Score 873.5; DP: 1; length 482.
Best Local Similarity: 39.5% Fred N= 37.5
Matches 187; Conservative: 91; Mismatches 187; Indels 1; Gaps 1.

QY 11 IDGQVYWRGMDIDVNPATEAVYISRIIDQGAEDARKAIDAERKQPEWALPAIERAS 70
  11:..... 111111 111111 111111 111111 111111 111111 111111
Db 15 INGEIMDANGNAGIAIVTNPANDKLGSPKMGAIIFPAIDANPAI PAMRAITAKREPAT 74
  15:..... 111111 111111 111111 111111 111111 111111 111111
QY 71 WLPKISAGIRFEASEISALIVEEERKTIQIAFVFAVFAIYIIVYMAFMAFVYDELISLD 130
  71:..... 111111 111111 111111 111111 111111 111111 111111
Db 75 ILKNMFMMEHODDLARLMTLEDOCKPLAEKKGISTASFTIEMFAEBKRIYGTIFGH 134
  75:..... 111111 111111 111111 111111 111111 111111 111111
QY 131 KPGENILLKFKALSVITGILPMNPEFLIAKMAFALDNIIVIPSEFTINMAIAFAK 190
  131:..... 111111 111111 111111 111111 111111 111111 111111
Db 135 QADKRLIVIKQPIGVTAATIPWNPFAAMITFKKAPALACQITWLRKASQTFPSALALAE 194
  135:..... 111111 111111 111111 111111 111111 111111 111111
QY 191 IYVFIQIPGCVNVLVLGGETVYGOELAGNPVAVNMTGVSVSGEIKMTAAKNITKYTL 250
  191:..... 111111 111111 111111 111111 111111 111111 111111
Db 195 LAIRACVAPGVNNTVTSNAGAVGNETSNPLVRLKLSFTGSTETGRQIMEDQAKDIKKSL 254
  195:..... 111111 111111 111111 111111 111111 111111 111111
QY 251 ELCKKAPATVMDADALELAVKAIYVSRVINSQVCNCAEVYVQKGIYEDPFVNRILAEAMQ 310
  251:..... 111111 111111 111111 111111 111111 111111 111111
Db 255 ELQGNAPFIYFDADILIKAVESGLASKRPMNGJTVVCANLTVYDGVDRPRAEKLIQVANS 314
  255:..... 111111 111111 111111 111111 111111 111111 111111
QY 311 AVQFGPARNDIANGPILNMAALERVQKARAVFEGAAVAVAGGKAVESKTYTPPTLL 370
  311:..... 111111 111111 111111 111111 111111 111111 111111
Db 315 KLHIGIGGLD-NGVTIGPLIDEKAAKVEEHIADLEKGAHVCGGKAHNRGGVFQPTLL 373
  315:..... 111111 111111 111111 111111 111111 111111 111111
QY 371 LDHROEIMSIMHEETGCPVLPVVAEDTLEDAISMANDSDYGLTSSITQNLINMAAKIKGL 430
  371:..... 111111 111111 111111 111111 111111 111111 111111
Db 374 VQVPAKAVKSEETFGCLAPLEFRFEDADVIAQANDTEFOLAAVFAARDLSVYFRGEL 433
  374:..... 111111 111111 111111 111111 111111 111111 111111
QY 431 KRGETYINRPNFAMQGFHAGHRKRSKGTGADGKRIHGYLQ 471
  431:..... 111111 111111 111111 111111 111111 111111 111111
Db 434 EYGVIGINTGIIISNEVAPGGIKASGLGREGSKGIDEDYLE 474
  434:..... 111111 111111 111111 111111 111111 111111 111111

RESULT 3
GABD_RHISN
ID GABD_RHISN STANDARD: PRT: 491 AA
AC P5653;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable succinate-semialdehyde dehydrogenase [NADP+] (EC 1.2.1.16)
DE (SSDH).
GN GABD OR Y4SL.
GN Rhizobium sp. (strain NGR234).
OS Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_Taxid=394;
RN 11
RP MEDLINE=67305956; PubMed=9163424;
RX Feidberg C.A., Fellay R., Baitoch A., Broughton W.T., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC 1- CATALYTIC ACTIVITY: Succinate semialdehyde + NAD(P)(+) + H(2)O -
CC succinate + NAD(P)H.
CC 1- PATHWAY: 4-aminobutyrate (GABA) degradation;
CC 1- SIMILARITY: RELATES TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL, institution -
CC the European Bioinformatics Institute. There are no restrictions on its use
CC use by non-profit institutions as long as its content is in 20 words
CC modified and this statement is not removed usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

Query Match	Best Local Similarity	33.1%	Score 810	DB 1	Length 491
Matches 159	Conservative 85	Mismatches 190	Indels 2	Gaps 2	
DR EMBL: AEO00096; AAR91449.1; -					
DR HSSP: P05031, 1CM3.					
DR IPIR: P05031, IPR02096; A3-b3y3-b3y3r.					
DR Pfam: PF00171, aligned. 1					
DR PROSITE: P50070; ALDEHYDE_DEHYD_CYS; FALSE_NEG.					
DR KMW PROSITE: P50067; ALDEHYDE_DEHYD_GLU; 1.					
DR Oxidoreductase; NADP; plasmid.					
FT NP_BIND 241 246 NADP (ADP PART) (BY SIMILARITY).					
FT ACT_SITE 263 263 BY SIMILARITY.					
FT ACT_SITE 297 297 BY SIMILARITY.					
SD SP09PFE 491 AA: 54253 MD: 67404 PF0410R14 C7064;					
OY 26 VNNPAREVATISRLPQOAGDADAKKALDAAEKQPEMELPALEASWILKISALIREKAE 85					
RB 37 VFNPSIGSELLAEVPIPMGSAALAHAALEKALAEPMWSGLIAAPSCILLMKHPFLHSID 96					
OY 86 ISATIVFEGGKTIQGLAEVAVATADITYMAFMAPYEGETIQSDPGENILFLPKALCV 145					
RB 97 IAAITLAFMGRKPIGFPAKSEFVQAIAAVIQWYAFAPATYGETISAPSTDPMLYIKQICV 156					
OY 146 TTGILFWNPEPFLIAKMAALLTNTNIVIKPSPTTNNALIAFAKTYNFGIDPQVFNIV 205					
DB 157 VCAITFWNPFASVAVAKISPAALACITVYLKFAEETPLVAMFMALAKIAFPQVYINIV 216					
OY 206 -LQGETVQGLAGNKKVMVAMTQSVSAGEKIMATAKNTTKVLELQSKAPADIVDA 264					
RB 217 YASPAITSPFICTNKKVKEITFTSTVQGLIMWQSTQIKRISFLQGNAPPIVDA 276					
OY 265 DELAAKATVDSKVINSGVQNCDAEKVYVQKSYLQFVNRLEAMQAVQGNPAERNDIA 324					
RB 277 ELDAAVVGAIDVAFPMAGTIVSANPIYVQSVVAEAFKTEFPVPLIKVAFEDPV-VA 335					
OY 325 MGPLINAALEIREKQAVARVEEGARVAFGGKAVNGKITYYPPILLLDVQEDMSIMEET 384					
DB 336 TGPLINDCAFKIKELHISNAVQKQAVPSGQNPRTGSSSTPEPPIVTVVSKTMPLAEET 395					
OY 385 FGLVLEVVAFETLEIALISMANSEUGLTSIYTLQNLVAMAKIIGLKEGYIINFEPTA 444					
RB 396 FGLIAPLILPHLEADAHVPEANITTYGLAAVYVASTLKRVWVVAALIPYCWVGINISMS 455					
OY 445 MGFSEHAWPKRSIGTALKEHGLHYLV 470					
DB 456 EAAPFGVGVKSGIGREGSGHGLEDTL 481					
RESULT 4					
DBAB_BACSU	STANDARD;	PRT;	490 AA.		
AC P71016;					
DT 01-NOV-1997 (Rel. 35, Created)					
DT 01-NOV-1997 (Rel. 35, Last sequence update)					
DT 15-JUN-2002 (Rel. 41, Last annotation update)					
DE Betaine aldehyde dehydrogenase (EC 1.2.2.18) (BADH).					
GN GBSA.					
OS Bacillus subtilis.					
OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.					
OX NCBI_TaxID:14233;					
RN [1]					
RP SEQUENCE FROM N.A. AND SP09PFE OF 1-25.					
RC STRAIN-168 / JH642;					
RX MEDLINE:96359364; PubMed-8752328;					
RA Boch J., Kempf B., Schmid R., Bremer E.;					
RT "Synthesis of the osmoprotectant glycine betaine in Bacillus					
RT subtilis: characterization of the gbsAB genes.";					
PL 1-1997; 1:483-151-529(1996)					
PN [2]					
RP SEQUENCE FROM N.A.					
RC STRAIN-168;					
RX MEDLINE:98044033; PubMed-9384377;					

120 6 ERK3V1KTAIELLRDLEELAELESLDTGKLTLESKADMDI LANNVQYAYGLADKGGEL 122

121 127 IQSDRGENTLILFKRALGVTTGILLPMPNPFLLARKKADALITGNTLVIKPSFTTNAL 186

122 125 ISSDILHDSBKSI LKEFLGVQGLTFWANNPLGLASMKIAPLALGNTLVMPKSPITPLTL 184

123 187 AFAFVIELLEIPGVNVLVSGEIVSGEFLAGNKKVMVAMTGSVSAAGEIMATANKIT 246

124 185 KVKIMPFPAVVRGVANVILVSGAVLVGRI LAVNKVGLI LSPGTLITLTKKIMAAAGNKK 244

125 247 KCLLEIGGAKPAIVMDADLELAVKAVIDSVINSGVNCMAEKRYVVGKIVYQVNRIL 306

126 245 KIALTEGGKRPNVLEKDAELEVAVQALNAVFPAGGVQVSGSKLLEVDALIHQFLAEVL 304

127 307 EAMQAVQFQPAERNDIANGPILINNAALEREQKVAARAEGARVAPGKAVG----KG 362

128 305 KKAKPIKIKGN-GHFAFTPSGILISFHAKEKRYVEIGLESAKLEFGSKPREPDELQNG 363

129 363 YVYPPFLLLVQKQMSIMHEETFGVLVVAFTLELAIATMANDSDYGLTSITVQNLNV 422

130 364 FFFVPEFNSQNSDMPLVQEFVFGVIVLVEFFSSSEFVIELANTITVGLAGAWSKDIEK 423

131 423 AMATIGLPGCEYINPEFTAMQGFHAGMKSLTGADKHGILHVTQVYV 476

132 424 GFVAAPLIPMTVWINPFFVFAQAPWGVKQSGFPELQFTGLEFTEVFNIV 477

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

5


```

DB 208 TGNVIVAKTASTSLMAIVSKYIPGAGIPPGVINIVSKGKVGATINNHKIKKAVAT 267
QY 229 GSVNAGKIMATAKNTKVCLELGAPAIIVMDDELAVKAIIVSKVINSOVNCA 488
DB 268 GSTATGRHIVOSAAAGIKKVTLELGGSPNIVPADLKAAYONILIGIYNSGEVCCAG 327
QY 289 ERVYVOKIYDQFENRIGEMOAVQFENRNDIAMGPLINAAALEREVOKARAVEG 348
DB 328 SRVVEESTYDKFIEEKKAASESKVADPPESTFVGAATSSMJLNKILKTYDLEKNEG 386
QY 449 ARVAGKAVHKGYYPPILLDVROKSTIMEETPGVLPVVAFTLEDAISMANDSD 408
DB 487 ATLLTGGKRLSSNGYFKPTVPGIVKEDMKIVKEELHGPVYVTKESADEVINMANDE 446
QY 409 YGLTSSITYTONLANYAKAIKIGKRGFTYINRPNFPMQGFH-----AGMPKSGTGADPK 463
DB 447 YGLAAGIHHSNINIALKAVADKAVNAGIWMINTYN-----DHHAVPFGFNASGLAKEMSV 501
QY 464 HGHAGYLOTOVV 475
DB 502 DALGNYLOVKAV 513

```

RESULT 7

```

DBAL_AGAB1 STANDARD: PRT; 500 AA.
AC 074187:
DE 16-OCT-2001 (Rel. 40, Created)
DE 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Aldehyde dehydrogenase (EC 1.2.1.3) (AlDH1).
GN ALDH1.
OS Agaricus bisporus (Common mushroom).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Agaricaceae; Agaricus.
OX NCBI_TaxID=5341;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Hort 839;
RA Schap P.L., Muller Y., Vissot J.;
RT Molecular structure and spatial expression of housekeeping genes in
mushrooms.
KL Submitted (Jun-1998) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC -1- PATHWAY: Ethanol utilization; second step.
CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isd.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y17825; CAA76875.1;
DB HSSB: P51977; IBSX.
DB InterPro: IP002086; Aldehyde_dehydr.
DB Pfam: PF00171; aldhc1.
DB PROSITE: PS00070; ALDEHYDE_DEHYDR_YS 1
DB PROSITE: PS00687; ALDEHYDE_DEHYDR_GLD 1.
KW oxidoreductase; NAD.
FT NP_BIND 246 251 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 269 269 BY SIMILARITY.
FT ACT_SITE 403 403 BY SIMILARITY.
SQ SEQUENCE 500 AA; 54195 MW; CAPCE5HR5085925 Ck0b4;

```

```

Query Match 41.4% Score 769; DB 1; Length 500;
Best Local Similarity 45.0%; Prod. No. 2.4e 44;
Matches 169; Conservative 104; Mismatches 198; Indels 12; Gaps 9;

```

2 SVVYDHMYIIVGQFVWIKGAWIIVVNPATFAVISTPAGQAFARAKAIIDAEKFA-GPEW 60

```

DB 17 STSINTGLFINGERVDVSKNTIIVVNPARGKILITISATRTADIDIAEAHAKAFETW 76
QY 61 EA-LPALEKASWLMKISAGIKERASELSALLIVEBKS LUOLAE-VEVAFTADYIDYMAW 118
DB 77 GLNCGSKRGDMYKIAQLMEKNIDIDALEADNCKETPLAMKSVLSISTIHVAGW 136
QY 119 AKHVEGELLSDRGENIILFKR--ALGVTTGILPNPFPFLIARKMAPALLTGNITVK 176
DB 137 AKKPNVYIETD---EKKLITYSHHPLGVVYKQIIPNNPILMIAMKIDHALTNGYIK 193
QY 177 PSEFTTNNAIAPAKIVNRTGIPPGVPMIYIGRGTWQGFACNPKVAVYSMGSSVAGK 236
DB 194 PSEFTPLALAKMCLALIDEAGFPFGVNVVYIGYSTTGOALSSHMKIDKVAFTSTLVGK 253
QY 237 IMATAAK-NITKYVLELGSKAPAIIVMDNATPLAVKAIIVDSVTNSGYNCAFEVYVK 295
DB 254 VMEFAAKSNIKNTVLEFGKSPVYITPDATFQSVWTAHGLFNNHQDCACTRIYVQE 313
QY 296 GTPQFVNPVIGFAMQAVQFQNPAPNPNTAMGPLINAAALPEYGVKAVAPFGAVAPVAG 355
DB 314 GIVYKPLKCTOKIKELKIDPRTG--IIVGIVYSIYDPTIMSYTESGRAGATVHVGG 372
QY 356 KAVGKGYPPPTLLLDVROKSTIMEETPGVLPVVAFTLEDAISMANDSDYGLTSS 415
DB 373 ERHNGEYFLOPTFTDTPDKIVKEELFGPGAVIKERDKEVYIKQANDSNYGLAAV 432
QY 416 YTONLANYAKAIKIGKRGFTYINREN-PEAQGFHAGMKKSGIGALKHGHAGYLOTOV 474
DB 433 FSDINNAITAIAPAKGTAWVNCANTIDAGVF-GGYKOSIGRIGETALHNTVYKA 491
QY 475 VYL 477
DB 492 VHV 494

```

RESULT 8

```

SSDH_RAT STANDARD: PRT; 488 AA.
AC P51650;
DE 01-OCT-1996 (Rel. 34, Created)
DE 01-OCT-1996 (Rel. 34, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Succinate semialdehyde dehydrogenase (EC 1.2.1.24) (MAD(+)-dependent
GN ALDH5A1 OR SSADH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (LINUS AND SHORT ISOPHENS), AND PAPPIAT SEQUENCE.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=95113870; PubMed=7814412;
RA Chambliss K L., Caudle P L., Hinson P P., Mowman C P.,
Slaughter C A., Jakobs C., Gibson K M.;
FT "Molecular cloning of the mature NAD(+) dependent succinate
semialdehyde dehydrogenase from rat and human. cDNA isolation,
FT evolutionary homology, and tissue expression."
RL J. Biol. Chem. 270:461-467(1995).
CC -1- CATALYTIC ACTIVITY: succinate semialdehyde + NAD(+) + H(2)O =
succinate + NADH.
CC -1- PATHWAY: 4-aminobutyrate (GABA) degradation.
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; a long form (shown here) and a
short form; may be produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: BRAIN, PANCREAS, HEART, LIVER, SKELETAL
MUSCLE, KIDNEY, LOWER IN SPLEEN, LUNG, KIDNEY AND TESTIS.
CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way

```


QY 74 KISAGIRERASELSALIVFEGCKIQOLAEVEVAFTADYIDYMAEWARVEGEFIQSDDPP 133
 Db 126 KWNLMONKDDARITAEKSKPLKEANGEILYSAPFLESEFARVYGGIHTPAKO 185
 QY 134 ENILFKRALGVTTGILPNWPFELIARKMAPALLTNGTIVIKPSEFTTNNAIAPAKIVD 193
 Db 186 RRALVLKQPIGVAAVITPNWPFISAMITRKVGGAALAGCTVVKVPAEDTFFSALALAEAS 245
 QY 194 EIGLPGEVENLY---LGRKETVQGLAENP/KVAMYSMTGSVSAGEKIMATAAKNITKVL 250
 Db 246 QAGTSGSVNVNTPCSEKNAKFEACITCDPLVSKISITFGSTTCKILLHHAANSKVRYSM 305
 QY 251 ELGGKAPAIMDDADELAVKALVDSRVINSQVCNCAERVVVQKGIYDQFVNRLGAMQ 310
 Db 306 ELGGLAPFVDSANVDQAVAGAMASKPRTQTCVCSNQFLVQGIHDAFVKAFEAAMK 365
 QY 311 A-VQGNPARNIDIAMPLINAAALFRVQKVARAVEEPARVAFGGKAVEGKGYYPPTL 369
 Db 366 KNLRVNGCFPG-FTQGPLINEKAVEKVKQVNDADVSKGATVVTGGKRHLQKGNFEPTL 424
 QY 370 LLDVPAFMSIMHETEGPVPVVAEDTLEDAISMANDSDYGLTSSIIYTNLNVAMKALIG 429
 Db 425 LCNVTQDMLCTHEETPGPLAPVKEDEEELATIANAADVGLAGYFSQDPAQIWRVAEQ 484
 QY 430 LKFCETVYINRENFEAMQGHAGWRKSGGIGGANGKHKHGLHGYLQTVV 475
 Db 485 LEVGMVGNELGLSSVECPFGVGKQSGLGREGSKYGDIEYLELKV 530

RESULT 10

DHAS_YEAST STANDARD; PRT: 519 AA.
 AC P40047;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Aldehyde dehydrogenase, mitochondrial precursor (EC 1.2.1.3).
 GN ALD5 OR ALD5 OR ALD3 OR YER073W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 NC Saccharomycetales; Saccharomycetaceae; Saccharomycetes
 NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97212813; PubMed=9059641;
 RA Wang X., Bai Y., Ni L., Weiner H.;
 RT "Saccharomyces cerevisiae aldehyde dehydrogenases: identification and
 expression.";
 RL Adv. Exp. Med. Biol. 414:277-280(1997)
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Dietrich F.S., Mulligan T.T., Hennessey K.M., Allen F., Araujo P.,
 RA Aviles E., Herron A., Brennan T., Carpenter J., Chen F., Cherry J.M.,
 RA Chung E., Duncan M., Guzman F., Hartwell G., Hunicke-Smith S.,
 RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
 RA Musedale D., Nakahara K., Namath A., Nargren P., Olfner P., Oh C.,
 RA Patel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
 RA Taylor P., Wei Y., Yellon M., Rotstein D., Davis R.W.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=98132377; PubMed=9473035;
 RA Wang X., Mann C.J., Bai Y., Ni L., Weiner H.;
 RT "Molecular cloning, characterization, and potential roles of cytosolic
 and mitochondrial aldehyde dehydrogenases in ethanol metabolism in
 Saccharomyces cerevisiae.";
 RL J. Bacteriol. 180:822-830(1998).
 CC 1- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
 CC 2- PATHWAY: Ethanol utilization; second step.
 CC 3- SUBCELLULAR LOCATION: Mitochondrial matrix (Potential).
 CC 4- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 CC

CC this SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch>)
 or send an email to license@isb-sib.ch.
 CC
 DR EMBL: U56605; AAB01220.1;
 DR EMRL: U48414; AAB44612.1;
 DR HSSP: P51977; IHX5.
 DR SGP: S0000875; ALD5.
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldedh; 1.
 DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GDO; 1.
 KW oxidoreductase; NAD; Mitochondrion; Transit peptide;
 TRANSIT 1 23 MITOCHONDRION (POTENTIAL).
 FT CHAIN 1 24 ALDEHYDE DEHYDROGENASE.
 FT NP_BIND 24 519 NAD (ADP PART) (BY SIMILARITY).
 FT ACT_SITE 287 287 BY SIMILARITY.
 FT ACT_SITE 321 321 BY SIMILARITY.
 FT CONFLICT 48 48 T -> I (IN REF. 2).
 FT CONFLICT 90 90 L -> A (IN REF. 2).
 FT CONFLICT 93 103 LLEPSVYKAG -> APTKWSIVPE (IN REF. 2).
 FT CONFLICT 410 410 F -> G (IN REF. 2).
 SQ SEQUENCE 519 AA; 56552 MW; 5D333F54477977FC CRC64;
 Query Match 30.7%; Score 750; DB 1; Length 519;
 Best local similarity 33.5%; Pred. No. 4644;
 Matches 169; Conservative 102; Mismatches 184; Indels 50; Gaps 9;
 QY 1 MSVPVQHP-----MYIKQFQVTKWKGIAHIIIVNVNATEAVISRIPIKQAFDAKKA 50
 Db 28 LRVPTILPMGFTYEQPTGLFTNGEEVAVSKQKTEFIVINSNEEKILIVYKAMEIIVDEAV 87
 QY 51 DAAERAQPEWEALPATERASWIKISAGIHERA-----SEISALIVEEGKIQ 98
 Db 88 AALKK-----LKKRSVYCR---AGVRAKALFNLAULVEKIQETLAAIESMUNGKSL 135
 QY 94 ELAEVFAVATAIYIYMAFMWAFKPEFELISLDEKFNILF KALGVTTIGILWNPFF 156
 Db 136 FCARGIVAVLSYKSGGWAIRIVGNVDT---GRNHTYSIKREPLAVVQIIPWNPFL 192
 QY 157 FLIAERKMAFALLIINTIVIKSEFTIRNAIAIAIVDEELIGVIVRELVIQGGTYSDEL 216
 Db 193 LMSWKIGIPALATGNTVILKPAETPLSALSLAQEGAGLPAVWNIIIPSGPVPVSGRL 252
 QY 217 ACNPVAMVSMLESVAPEFIMATAAKNIDFWTLFGKALAVMGAALFLAYKAIVDS 276
 Db 253 SAIPIVKRIAPTGSTATGHHIMKVAATVVKVKTLEGGKSNIVAIADGLVAVKNIAPI 312
 QY 277 RVINSQVCNVAEPVVVAPVLYTGPNRLGAMQAVQGNFAEKNDIAMHPLINAAALER 336
 Db 313 IFYNSCEVCCASRIYIQVTVEEVLOKLKAYTESLKWDDPP EEPVCACTSNKQLHR 371
 QY 347 VEGKVAFAVEFGARVAFGKAVEGKGYYPPTLIIIVRQEMSIMHETFGVPLVVAFTD 406
 Db 372 ILDYVEVAKSEGARLVTGARHGSKGYFVKVTVFAVKEIMPFVKEEVFGPITVTSKEST 431
 QY 397 LEDAISMANDSIYGLTSSIIYTNLNVAMKALIKLKGKPGTYINRNFNEMAKGFEH 451
 Db 432 VDEIVAMANDSOYGLAAGIHTNDINKAVIVSKRKAKTIVWNIYN NFRNVVTFGG 486
 QY 452 WRKSGIGGADKHKHGLHGYLQTVV 475
 Db 487 FGSGIGRGMGFAALSNVTQTKSV 510
 RESULT 11
 DHAM_LEITA
 ID DHAM_LEITA
 AC Q25417; PRT: 498 AA.

DT 01-NOV-1997 (Rel 35, Created)
 DT 01-NOV-1997 (Rel 35, Last sequence update)
 DT 15-JUN-2002 (Rel 41, Last annotation update)
 DE Aldehyde dehydrogenase, mitochondrial precursor (EC 1.2.1.3) (ALDH
 DE class 2) (P51).
 GN ALDH2.
 OS Leishmania tarentolae (Sautoleishmania tarentolae).
 OC Eukaryota, Eulicostoa, Kinetoplastida, Trypanosomatidae, Leishmania.
 OX NCBI_TaxID=5689;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 10-31.
 RC STRAIN=UC;
 RX MEDLINE=95356798; PubMed=7630384;
 RA Bringaard F., Paris M., Zen K.H., Simpson L.;
 RT "Characterization of two nuclear-encoded protein components of
 RT mitochondrial ribonucleoprotein complexes from Leishmania
 RT tarentolae." Parasitol. 71:65-79(1995).
 RL MOL. Biochem. Parasitol. 71:65-79(1995).
 CC -!- FUNCTION: COULD HAVE A RNA-BINDING ACTIVITY IN ADDITION OF ITS
 CC CATALYTIC ROLE.
 CC -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
 CC -!- PATHWAY: Ethanol utilization; second step.
 CC -!- SUBCELLULAR LOCATION: Mitochondrial.
 CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC
 CC EMBL: Z31698; CAAB3503.1; -;
 CC HSSP: P05091; ICW3.
 CC InterPro: IPR002086; Aldehyde_dehydr.
 CC Pfam: PF00171; aldedh.1
 CC PROSITE: PS00070; ALDEHYDE_DEHYDR_GLU; 1.
 CC PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
 CC Oxidoreductase; NAD; Mitochondrion; Transit peptide; RNA-binding.
 KW TRANSIT 1 9 MITOCHONDRION
 FT CHAIN 10 498 ALDEHYDE DEHYDROGENASE.
 FT NP_BIND 242 247 NAD (ADP PART) (BY SIMILARITY).
 FT ACT_SITE 265 265 BY SIMILARITY.
 FT ACT_SITE 299 299 BY SIMILARITY.
 FT ACT_SITE 498 498 BY SIMILARITY.
 SQ SEQUENCE 498 AA, 618055F6ED5547EC CRC64;
 Query Match 29.8%; Score 728; DB 1; Length 498;
 Best Local Similarity 35.1%; Pred. No. 1.3e-41;
 Matches 168, Conservative 99, Mismatches 204, Indels 8, Gaps 7;
 QY 5 VQHPMY IUGQFVTRKGLAWIDVYNPAEAVISKIPDQDAEDARKALDAAEKAPQWPAALP 64
 DE 18 IQEKLIINKEFVPAVSQKTFVYNPADEKVIANVAEAKADVLAVAKAARHAFESERMTD 77
 QY 65 AIERASWLRKISAGIPERASEISALIVEREGKIQOLA-EVFAFTADYIDYMAEWAPRYE 123
 DE 78 QMPNDUMIRIADILEKNSKEMAAEISLNGKPYEVALNVVALSVCEPRYCAGIADKVN 137
 QY 124 GEITQSPRPQFNILLKRP-ALGVTTGILPWNPFFFLIARKMAFALITGNTIVIKRSEFTT 182
 DE 138 GTV--PPRSNGFLGKVPQPGVGGIIPWNPFLIMAAFKLSFALAMNTVVLKAEVTP 195
 QY 183 NNATFAKIVDEICLGRVFNIVRGFTVGGFGLACNPKVAMVWFGVSAGEKIMATAA 242
 DE 196 LTAVRLGEMVMEAGYPDGVNLLPGFGATAGSEIAHMDVDVRIAFVTSVAVGHVQMMAA 255
 QY 243 K-NITKYLELGGKAFATVMDAGLELAVKAVUSKVINSGAVQWNAERVVYKQKLYDF 401
 DE 256 ETNLKYSLELGGKSAIVCEADADGAEAAEAVATIKVYFTNGVUCVASSKIYVHESYDFE 315
 QY 302 VNRLEAMAVVQGNPAERNDIAMKPLINAAALEKVFQKVAFAVEEGAFVAFGKAVBGR 361
 DE 361

DE 316 VSLRKNAEAPKVC-PCNDTGNMGPFLVSKQHEPVGVGVTFDGVVAGATVVTGCKKIGUK 374
 QY 362 GYYPPPLILLDVRQMSIMHEEIFGVLPVAVAFUFLDAISMANISDFGLTSSITYNLIN 421
 DE 375 GYFVQPTIFSDKREDFPKFPGFVTCVMKYGFMDFVVRPANDSISVSLAAGICTPSMD 434
 QY 422 VAMKAKGLKFGFGEIYNP-ENFEAMGSHASWPKSGIGAGAKHGIHGVLTQIVVYLOS 479
 DE 435 TALRYSTYLNAGTIVWNTNNTCPSPMPF-GGFKQSGIGBELGKFKVVDVMTPEKAIHFAS 492
 RESULT 12
 DHAL-EMENI
 IL QHAL-EMENI STANLUKU; PRI; 497 AA.
 AC P08157;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Aldehyde dehydrogenase (EC 1.2.1.3) (ALDH).
 GN ALDA OR ASPA.
 OS Emericella nidulans (Aspergillus nidulans).
 OC Eukaryota, Fungi, Ascomycota, Pezizomycotina, Eurotiomycetes,
 OC Eurotiiales, Trichocomaceae, Emericella.
 OX NCBI_TaxID=5072;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=8724800; PubMed=3024652;
 RA Pickert M., Gwynne B.L., Rusten P., Elliott P., Davies P.W.,
 RA Lockington R.A., Scanzocchie C., Sealy-Lewis H.M.;
 RT "Cloning and characterization of the alda gene of Aspergillus
 RT nidulans.";
 RL Gene 51:217-226(1987).
 CC -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
 CC -!- PATHWAY: Ethanol utilization; second step.
 CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC
 CC EMBL: M16197; AAA33293.1; -;
 CC FIP; A29055; A29055.
 CC HSSP: P05091; ICW3.
 CC InterPro: IPR002086; Aldehyde_dehydr.
 CC Pfam: PF00171; aldedh.1
 CC PROSITE: PS00070; ALDEHYDE_DEHYDR_GLU; 1.
 CC PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
 CC Oxidoreductase; NAD.
 FT NP_BIND 241 245 NAD (ADP PART) (BY SIMILARITY).
 FT ACT_SITE 264 264 BY SIMILARITY.
 FT ACT_SITE 298 298 BY SIMILARITY.
 SQ SEQUENCE 497 AA, 54088 MW, 171FCFAQ93E571B CRC64;
 Query Match 29.7%; Score 726; DB 1; Length 497;
 Best Local Similarity 33.7%; Pred. No. 1.8e-41;
 Matches 159, Conservative 98, Mismatches 207, Indels 8, Gaps 5;
 QY 9 MYIDGFVTRKGLAWIDVYNPAEAVISKIPDQDAEDARKALDAAEKAPQWPAALP 67
 DE 21 LFINNEFVKQVEKTKTQVINFNSKRVTSVHEATEKDVAVAAARAAAFEGFWRVTPSE 80
 QY 68 KASWLRKISAGIKERASEISALIVEREGKIQOLA-EVFAFTADYIDYMAEWARKYEGEII 127
 DE 81 PGLINKLAKMEKPIHITAAIESLNGKAFIMAKVILANSIGGLKYYAGWAKIHQGTI 140
 QY 128 QSDRPGENILLFKR--ALGVTTGILPWNPFFFLIARKMAPALLTONTIVIKRSEFTTNA 185
 DE 141 DTN---PETLYTIKHEPVGVCGIIPWNPFLIMAAFKLSFALAMNTVVLKAEVTP 197

[illegible]

RESULT 13

1D	DHA6_HUMAN	STANDARD;	PRT;	512 AA.
AC	P47895;			
DT	01-FEB-1996	(Rel. 33, Created)		
DT	01-FEB-1996	(Rel. 33, Last sequence update)		
DT	15-JUN-2002	(Rel. 41, Last annotation update)		
DE	Aldehyde dehydrogenase 6 (EC 1.2.1.5).			
DE	Aldehyde dehydrogenase 6 (EC 1.2.1.5).			
GN	ALDH1A3 OR ALDH6.			
OS	Homo sapiens (Human).			
OC	Pekaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Salivary gland;			
RA	MEDLINE=95213025; PubMed=7698756;			
RX	Hsu I.-C., Chang W.-C., Hiraoka L., Hsieh C.-L.;			
RT	*Molecular cloning, genomic organization, and chromosomal			
RT	localization of an additional human aldehyde dehydrogenase gene,			
RT	ALDH6.*;			
RL	Genomics 24:333-341(1994).			
CC	-1- CATALYTIC ACTIVITY: An aldehyde + NAD(P)(+) + H(2)O -> an acid +			
CC	NAD(P)H.			
CC	-1- PATHWAY: Ethanol utilization; second step.			
CC	-1- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVELS IN MANY TISSUES AND AT			
CC	HIGHER LEVELS IN SALIVARY GLAND, STOMACH, AND KIDNEY.			
CC	-1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.			


```
FT ACT_SITE 257 257 BY SIMILARITY
FT ACT_SITE 291 291 BY SIMILARITY
FT CONFLICT 424 424 S -> F (IN REF. 2)
SQ SEQUENCE 497 AA: 54270 MW: 55088240635B22F CRC64:

Query Match
Best Local Similarity 34.2%; Score 718.5; Dh 1; Length 497;
Matches 162; Conservative 104; Mismatches 203; Indels 19; Gaps 8;

QY 1 MSVPV-CHPMYIDGQFTWTRGDA--WIDVNPATEAVISRIPTDGOAEDARKAIDAAPRA 56
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 1 MAEPFAPARQLFIDGE---WREPTKKNRPDVINPSTEEIIGDIPAATAEDVEVAVVAARRA 57
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 57 --OPEWALPAIERASWLKKISAGIRERASEISALIVERGGKIQOOLAEVEVAFTADYIDY 114
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 58 FRNNWSATSCAHRATYLRALIAAKITEKKDHFVKLETIDSGAPFDEAVLDDDDVASCFEY 117
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 115 MAEWARKYEGE-----IIQSDRPGENILLEKRALGVTTGILPWNEPFFELIARKMAPALLT 169
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 118 FAGQAFALDGKQKAPVTLPMERFKSHVL--ROPFGVGLISPNWYPLLIMATWKIAPALAA 175
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 170 GNTIVIKPSEFTTNNIAIAFAKIVDEICLPRGVFNIVLGRGETVGGELAGNPKVAMVMTG 229
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 176 GCTAVLKPSLASVTCLERGEVGCNEVGLPPGVNIIITGLGPDAGAPIVSHPDVDKIAFTG 235
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 230 SVSAGEKIMATAAKNITKVCLLELGGKAPAIVMDDADLLELVAKAIYDSRVINSGOVCNAE 289
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 236 SSATGSKVMASAAQLVKPVTLLELCKSPVIVFEDVDIKVVPWTIFGGEFTNGGICGATS 295
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 290 KVVYKGIYDQFVNRIGEAQVQFONPAERNLIMAGPLINAAALERVEQKVARAVEGA 349
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 296 RLLVHESIAAEFVDKLVKWKTKIKISDPPEEG-CRLGPVISKGOYDKIMKFISTAKSSEA 354
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 350 RVAFGGKAVE--CKGYYYPTLLLOVROEMSIMHEETGCVLPVVAFTDLEDAISMANDS 407
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 355 TILYGGSRPHLKKGYIEPTIVTDISTSMQIWKEEVFGPVLCKVTFSSSEDEAIALANDT 414
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 408 DYGITSSIYTONLNVAMKAIKGLKEGTYINRENFEAMQGFHAGWRKSGIGGADCKHGLH 467
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 415 EYGLAAVFSNDLRCERITKALEVGAVVWNC:SQPCFVQAPWGGIKRSCFCGRELGEWGIC 474
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 468 GYLQTOVV 475
DQ : : : : :
Db : : : : :

QY 475 NYLNIKQV 482
```

Search completed: June 24, 2003, 10:14:29
Job time : 10.6475 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:08:15 ; Search time: 46.606 Seconds

(without alignments)
2862168 Million col: updates/sec

Title: US-09-830-751-6

Perfect score: 2445

Sequence: 1 MSVPVQHPYIDGQVPTWKG ADRKHGIRVYIQVQVVLDS 474

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 206047115

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SPTREMBL_21.*
- 2: sp_archaea.*
- 3: sp_bacteria.*
- 4: sp_fungi.*
- 5: sp_human.*
- 6: sp_invertebrate.*
- 7: sp_mammal.*
- 8: sp_mhc.*
- 9: sp_ordanelle.*
- 10: sp_phase.*
- 11: sp_plant.*
- 12: sp_rodent.*
- 13: sp_virus.*
- 14: sp_vertebrate.*
- 15: sp_unclassified.*
- 16: sp_rv.*
- 17: sp_bacteriap.*
- 18: sp_archaea.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	2426	99.2	474	16	Q9K901
2	1557	63.7	480	16	Q9JXMH
3	1552	63.5	480	16	Q9JW97
4	1279.5	52.3	393	16	Q9P119
5	1066.5	43.6	486	2	Q93088
6	903	36.9	493	17	Q9H001
7	892	36.9	493	17	Q979W0
8	894	36.5	492	16	Q9PKF1
9	879.5	36.0	483	16	Q916M5
10	872.5	35.7	497	16	Q9Y3P2
11	870.5	35.6	482	16	Q9X950
12	868.5	35.5	487	16	Q9CMM2
13	858.5	35.1	492	16	Q9Z4P8
14	845.5	34.6	475	16	Q9K7P5
15	840.5	34.4	484	16	Q9Z1P2
16	836	34.2	493	16	Q9ZV65

17	835	34.1	468	16	Q9KE63
18	832	34.0	486	16	Q9CKS0
19	831.5	34.0	485	16	Q9AL54
20	827	33.8	428	16	Q9SEK4
21	822	33.6	482	16	Q9A3R2
22	821.5	33.6	487	16	Q9Y157
23	814.5	33.3	484	16	Q9ZV70
24	814.5	33.3	509	10	Q9SAK4
25	813.5	33.3	485	16	Q98711
26	811.5	33.2	486	16	Q9YAL4
27	809	33.1	488	16	Q9ZDA3
28	804	32.9	495	16	Q9R827
29	803	32.8	482	16	Q94438
30	801.5	32.8	474	2	Q948F1
31	798.5	32.6	486	16	Q9U4H1
32	797.5	32.6	478	17	Q9ZV31
33	797	32.6	509	5	Q9VBP6
34	796.5	32.6	484	2	Q9VPD7
35	795.5	32.5	489	16	Q9UKH2
36	795.5	32.5	490	16	Q9Z2L8
37	795	32.5	494	16	Q9ZXX0
38	794	32.5	497	16	Q98AY0
39	793.5	32.4	493	16	Q9XPW7
40	791.5	32.4	487	16	Q9X946
41	785.5	32.1	484	2	Q9WMD5
42	784.5	32.1	484	16	Q9U852
43	783.5	32.0	553	10	Q91R16
44	782.5	32.0	474	16	Q9X965
45	781.5	32.0	477	16	Q9JTN7

ALIGNMENTS

RESULT 1

Q9X901	Q9X901	PEPELINAPRY:	PRT:	479 AA.
AC	Q9X901:			
DT	01-MAP-2002 (TREMREL: 20, Created)			
DT	01-MAP-2002 (TREMREL: 20, Last sequence update)			
DT	01-JUN-2002 (TREMREL: 21, Last annotation update)			
DE	Aldehyde dehydrogenase, NAD-linked.			
GN	ALDH4 OR 22306 OR ECS-021.			
OS	Escherichia coli O157:H7.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Escherichia.			
OX	NCBI_TaxID=83334;			
RN	[1]			
SP	SEQUENCE FROM N.A.			
STRAIN	O157:H7 / F0333 / ATCC 700927;			
XX	MEDLINE=21074935; PubMed=11206551;			
RA	Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,			
RA	Rose D.T., Mayhew G.F., Evans P.S., Gregor T., Kirkpatrick H.A.,			
RA	Probst G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,			
RA	Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,			
RA	Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,			
RA	Welch R.A., Blattner F.R.;			
RA	Nature 409:529-533(2001).			
FL	"Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";			
FL	Nature 409:529-533(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
FX	STRAIN=O157:H7 / H100_0509952;			
XX	MEDLINE=21157331; PubMed=11258796;			
RA	Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,			
RA	Han C.-G., Ohnishi E., Nakayama K., Murata T., Tanaka M., Tobe T.,			
RA	Jida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,			
RA	Kuhara S., Shiba T., Hattori M., Shinozuka K.,			
ET	"Complete genome sequence of enterohaemorrhagic Escherichia coli			
ET	O157:H7 and genome comparison with a laboratory strain K12";			
FL	UNA Res. 8:11-22(2001).			
FL	EMBL: AB005964; AA056490.1; ..			
DE	EMBL: AB005964; BAB35444.1; ..			
DE	Interpro: IIR02006; Alchyd_dehydr.			

```

DR pfam: PF00171; aldedh: 1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 479 AA; 52238 MW; D49819FA02E5C7CF CRC64.

Query Match
Best Local Similarity 99.2%; Score 2426; DB 16; Length 479;
Matches 476; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSVPVQHPMYIDGQFVTRGDAWLDVYNVPAEAVISKIPGQAEDAKKALDAERAQPEW 60
DB 1 MSVPVQHPMYIDGQFVTRGDAWLDVYNVPAEAVISKIPGQAEDAKKALDAERAQPEW 60
QY 61 EALPAIERASLWKISAGIRERASEISALIVEEGGKIQOLAEVEVAFTADYIDYAEWAR 120
DB 61 EALPAIERASLWKISAGIRERASEISALIVEEGGKIQOLAEVEVAFTADYIDYAEWAR 120
QY 121 RYEGEITQSDRPGENILLFKRALGVTTGILPWNPFFLIARKMAPALLTGNTIVIKPSEF 180
DB 121 RYEGEITQSDRPGENILLFKRALGVTTGILPWNPFFLIARKMAPALLTGNTIVIKPSEF 180
QY 181 TTNNAIAFAKIVDEIGILPRGVNVLVGRGETVGOELAGNPKVAMVSWTGSVAGEKIMAT 240
DB 181 TTNNAIAFAKIVDEIGILPRGVNVLVGRGETVGOELAGNPKVAMVSWTGSVAGEKIMAT 240
QY 241 AAKNITKVCLELGGKAPAIIVMDADLELAVKAIVDSKVINSGOVNCNAEKVVYQGIYDO 300
DB 241 AAKNITKVCLELGGKAPAIIVMDADLELAVKAIVDSKVINSGOVNCNAEKVVYQGIYDO 300
QY 301 FVNRLGAMQAVQGNPAENDTAMGPLINAAALERVEQKAVAFEGARVAFGGKAVBG 360
DB 301 FVNRLGAMQAVQGNPAENDTAMGPLINAAALERVEQKAVAFEGARVAFGGKAVBG 360
QY 361 KGYTYPPTLLDVRQEMSIMHEETFGPVLVPAVADTLEDAISMANDSDGLTSSIVTQNL 420
DB 361 KGYTYPPTLLDVRQEMSIMHEETFGPVLVPAVADTLEDAISMANDSDGLTSSIVTQNL 420
QY 421 NVAMKAIKGLKFGETYINRENFEAMOGFHAGWPKSGIGGADGKHGLHYLOTVVYLOS 479
DB 421 NVAMKAIKGLKFGETYINRENFEAMOGFHAGWPKSGIGGADGKHGLHYLOTVVYLOS 479

RESULT 2
Q9JW97
ID Q9JW97 PRELIMINARY; PRT; 480 AA.
AC Q9JW97;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Aldehyde dehydrogenase A.
GN NMB1968.
OS Neisseria meningitidis (serogroup B).
OC Bacteria, Proteobacteria, beta subdivision, Neisseriaceae, Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickoy E.K.,
RA Hatt D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Clifton H., Clark F.R.,
RA Cotton M.D., Utterback T.N., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masignani V., Pizzi M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappaport R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
MC58."
RL Science 287:1809-1815(2000).
DR EMBL: AE002544; AAP42297.1;
DR HSSP: P51977; 1BXS.
DR TIGR: NMH1968; 1.

DR InterPro: IPR002086; Aldehyde dehydr.
DR pfam: PF00171; aldedh: 1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 480 AA; 52256 MW; 45A72F5D1241F107 CRC64.

Query Match
Best Local Similarity 64.3%; Score 1507; DB 16; Length 480;
Matches 301; Conservative 64; Mismatches 110; Indels 2; Gaps 1;

QY 6 QHMYIDQGVTRGDAWLDVYNVPAEAVISKIPGQAEDAKKALDAERAQPEWALJA 65
DB 1 QHMYIDQGVTRGDAWLDVYNVPAEAVISKIPGQAEDAKKALDAERAQPEWALJA 65
QY 3 QHMYIDQGVTRGDAWLDVYNVPAEAVISKIPGQAEDAKKALDAERAQPEWALJA 62
DB 3 QHMYIDQGVTRGDAWLDVYNVPAEAVISKIPGQAEDAKKALDAERAQPEWALJA 62
QY 66 IERASLWKISAGIRERASEISALIVEEGGKIQOLAEVEVAFTADYIDYAEWAR 125
DB 66 IERASLWKISAGIRERASEISALIVEEGGKIQOLAEVEVAFTADYIDYAEWAR 125
QY 63 VERGAYLRKIAQGIHRADELDTTIVAEGGKIKDLARVEMFTADYIDYAEWAR 122
DB 63 VERGAYLRKIAQGIHRADELDTTIVAEGGKIKDLARVEMFTADYIDYAEWAR 122
QY 126 IQSDRPGENILLFKRALGVTTGILPWNPFFLIARKMAPALLTGNTIVIKPSEFTINNA 185
DB 126 IQSDRPGENILLFKRALGVTTGILPWNPFFLIARKMAPALLTGNTIVIKPSEFTINNA 185
QY 123 IQSDRPGENILLFKRALGVTTGILPWNPFFLIARKMAPALLTGNTIVIKPSEFTINNA 182
DB 123 IQSDRPGENILLFKRALGVTTGILPWNPFFLIARKMAPALLTGNTIVIKPSEFTINNA 182
QY 186 IYAKIVDEIGILPRGVNVLVGRGETVGOELAGNPKVAMVSWTGSVAGEKIMAT 245
DB 186 IYAKIVDEIGILPRGVNVLVGRGETVGOELAGNPKVAMVSWTGSVAGEKIMAT 245
QY 183 HIFAEIVDAVGLPAGVNVVNGFGAEIGNALSARHPQVDMVSLTGSVAGEKIMAT 242
DB 183 HIFAEIVDAVGLPAGVNVVNGFGAEIGNALSARHPQVDMVSLTGSVAGEKIMAT 242
QY 246 TKVLELGGKAPAIIVMDADLELAVKAIVDSKVINSGOVNCNAEKVVYQGIYDO 405
DB 246 TKVLELGGKAPAIIVMDADLELAVKAIVDSKVINSGOVNCNAEKVVYQGIYDO 405
QY 243 IKVLELGGKAPAIIVMDADLELAVKAIVDSKVINSGOVNCNAEKVVYQGIYDO 402
DB 243 IKVLELGGKAPAIIVMDADLELAVKAIVDSKVINSGOVNCNAEKVVYQGIYDO 402
QY 406 GEAMQAVQGNPAENDTAMGPLINAAALERVEQKAVAFEGARVAFGGKAVBG 464
DB 406 GEAMQAVQGNPAENDTAMGPLINAAALERVEQKAVAFEGARVAFGGKAVBG 464
QY 363 TAAMKGVYCNPAEAGALEMHLTEERAVAVAFKVEKAVKQAKLWACKKAKG 462
DB 363 TAAMKGVYCNPAEAGALEMHLTEERAVAVAFKVEKAVKQAKLWACKKAKG 462
QY 364 YYPPTLLDVRQEMSIMHEETFGPVLVPAVADTLEDAISMANDSDGLTSSIVTQNL 424
DB 364 YYPPTLLDVRQEMSIMHEETFGPVLVPAVADTLEDAISMANDSDGLTSSIVTQNL 424
QY 363 FFEPTLLDVRQEMSIMHEETFGPVLVPAVADTLEDAISMANDSDGLTSSIVTQNL 422
DB 363 FFEPTLLDVRQEMSIMHEETFGPVLVPAVADTLEDAISMANDSDGLTSSIVTQNL 422
QY 424 MKAIKGLKFGETYINRENFEAMOGFHAGWPKSGIGGADGKHGLHYLOTVVYLOS 479
DB 424 MKAIKGLKFGETYINRENFEAMOGFHAGWPKSGIGGADGKHGLHYLOTVVYLOS 479

SEQUENCE FROM N.A.
STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
MEDLINE=2022556; PubMed=10761919;
Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
Klee S.R., Morelli G., Basham D., Brown D., Chillingworth J.,
Davies K.M., Davis R., DeLisle K., Holtwell J., Hamlin R., Hargrett-
Nelson K., Leather S., Mouton R., Mungall K., Quail M.A.,
Kajander M.A., Parkhill J., Parkhill J., Parkhill J.,
Whitehead S., Spratt B.G., Bartell R.G.;
"Complete DNA sequence of a serogroup A strain of Neisseria
meningitidis Z2491."
Nature 404:502-506(2000).
EMBL: AL162754; CAH83774.1;
HSSP: P51977; 1BXS.
InterPro: IPR002086; Aldehyde dehydr.
pfam: PF00171; aldedh: 1.

```

```

DR PROSITE: PS00076, ALDEHYDE_DEHYDR_GLY; UNKN-WN_1;
DR PROSITE: PS00687, ALDEHYDE_DEHYDR_GLU; UNKNOWN_1;
KW Oxidoreductase, Complete Proteome;
SQ SEQUENCE 480 AA; 5255 MW; 1508407676347609 CRC64;

Query Match
Best Local Similarity 63.5%; Score 1552; DB 16; Length 480;
Matches 300; Conservative 64; Mismatches 110; Indels 2; Gaps 1;

QY 6 QHPMYIDGQFVTRGDADVDVNNPATEVATISRIIPDQQAEDARKAIDAAEQAPWELPA 65
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 3 QIANYINGRFENDENGWRNYNTEETAEAREPKGKADVDRAVAARAQAQAPWELPA 62
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 66 IEKASWLEKISAGIPEPASEYISALIVIEGSKLQCLAEVEVAFTADYIDYMAEWARVEGE 125
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 63 VEPQAYLPTAGIPEPARELDTDTYVAPGCKTKPLAPVEVFTADYLDYQAEWARVEGE 122
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 126 ITOSDRPKNILLKRAIGVTTGILPNWPFELIARKMAPALLTNGTIVIKPSEFTTNN 185
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 123 ITQSDPPENITLLEKPTDVGAGILPNWPFELIARKMAPALLTNGTIVIKPSSVTTFIN 182
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 186 IAFKIVDEIGLPGCVENLVLRGETVGOELAGNPKVAMVSMTGSVSAGEKIMATAAKNI 245
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 183 HIFAEIVDAVGLPAGVFNWNGPCAEIGNALSAHPQVMVSLTGSVEAGPQVWEAASANI 242
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 246 TKVCELGSKAPAIYMGDADLELAVKAIIVISKVINSGLVGNCAEKVYVVGKGIYDQFVNRL 305
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 243 TKVSELGSKAPAIYVLEKADILAVKSTLASPVNTQICNCAFFVYVHSSLSKDALEKM 302
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 306 GEMAOVQGNP--AERNDIANGPLINNAALERYEQKVARABEGAPVAGGKAVGKGY 363
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 303 TAAMKGVRYGNPAEAGALEMGLIEERAVKAVKVERAVKQAGLKGCGGRAREGRY 362
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 364 YYPPTLLLDVROEMSIMHEETFGVPLPVAFDTLEDALSMANDSDGLTSSIVYTONLVA 423
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 363 FEPTLLDTONSDIMKEETFGVPLPVSAFDTLDOVIALANDCEFGLTSSVYTNLNEA 422
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 424 MFAIKGLKFGETYINRENFEMAQPHAGWKSSTGTAAGKSHGLHGYLQTAQVYVLYQS 479
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 423 FYVTRFLQFGETYINRENFEMAQPHAGWKSSTGTAAGKSHGLHGLFEVILQVYVLYET 478
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 4
Q9P119 PRELIMINARY; PRT; 393 AA.
ID Q9P119
AC Q9P119;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Aldehyde dehydrogenase C-terminus (EC 1.2.1.22).
GN ALD' OR CJO490.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria, epsilon subdivision, Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC SPRAIN=NCr 11168;
PX MEDLINE=26150412; PubMed=1688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley T.M., Chircher C.,
RA Basham D., Chillingworth T., Davies K.M., Fellwell T., Holtroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallon M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences."
PL Nature 403:665-668(2000).
DR EMBL: AL139675; CAB75128.1;
DR HSP: P51977; IBOX.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aided. 1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; UNKN-WN_1.
KW Complete proteome.

```

Db 55 AERAKQKQKVPARKRADLRTTAAETKANKNOLAELIVREQGKLLAVAKMEVEVTASFI 114
 QY 113 DYMAEWARRYEGEIIQSDRGENTILLPKRALGVTTGILPWNFFFLIAKKMAPALLTGNF 172
 Db 115 EYACDGAKEGDIIPSDNPAEHIIMHKIPRGVVAITAMNFFPLAIGRKLGPALVAGNS 174
 QY 173 IVIKPSEFTTNNALAFKIVDEIGLPRGVNVLGRGCTVGQELAGNPKVAMVSMTGWS 232
 Db 175 VVLKPTQPTPLATLEIGNANKVGLPKGLNLTGGVSLGNALVAMPITKVVSMTGSTP 234
 QY 233 ACEPKIMATAAKNTKVCLELGGKAPALVMDADLELAVKAIIVSRVINSQVNCNCAERY 292
 Db 235 AGGGLFRTAENLHVGLLELGGKAPCIVFDADLEAVEGAFHSRFDNGGVCUTSNERLY 294
 QY 293 VOKGIYQDFVNRKLGCAEQVQGNPAERNDIANGPILINAAALERVEQKVARAVEGARVA 352
 Db 295 VHEISYNEFERFERMEKYKGLKLGNDPDA-ATTIGPKVNAKEVAHMEHLVTKSVEGATVA 353
 QY 353 FGKRAVEG- - - KGVYYPPTLLLDVROEMSIHHEETFGVPLVPVAFDTLEDAISMANDSD 408
 Db 354 IGGKKPOGKAFAKEGHWEPTLLTDVKONMTIVHEESFGPLVPVKESEFNEVIGYANDCE 413
 QY 409 YGLTSSYTONLVAMKAIKGLKFGETYINRENFEAMQGFHACWRKSGIGGALGRKHGJHG 468
 Db 414 YGLAAVFTDMNKMIMRLNDELEFGEIYINRGHGEHQHGFHNGYKLSGTGGEDSKYGFQ 473
 QY 469 YLQTVVYLO 478
 Db 474 YMEKKTFTYK 483

RESULT 6

Q9HK01 ID Q9HK01 PRELIMINARY; PRT: 493 AA.
 AC Q9HK01: 01-MAR-2001 (TrEMBLrel. 16, created)
 DT 01-MAR-2001 (TrEMBLrel. 16, last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)
 DE Probable aldehyde dehydrogenase.
 GN TA0809.
 OS Thermoplasma acidophilum.
 OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatiales;
 OC Thermoplasmataceae; Thermoplasma.
 OX NCBI_TaxID=2403;
 RN [1]
 RP SEQUENCE FROM N. A.
 RC STRAIN=DSM 1728;
 RX MEDLINE=20479972; PubMed-11079001;
 RA Ruepp A., Graml W., Santos-Martinez M.-I., Korotkova Y. P., Volker G.,
 RA Newes H. W., Frishman D., Stocker S., Lupas A. N., Baumgaertel W.;
 RA "The genome sequence of the thermoacidophilic scavenger Thermoplasma
 acidophilum.";
 RT Nature 407:508-513(2000).
 RL EMBL: AL445065; CAC11938.1; .
 DR HSSP: P05091; ICW3.
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldehyd_1.
 DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 493 AA; 54782 MW; 2F6A79DD96BE7FF5 CRC64;

Query Match 46.9%; Score 903; DB 17; Length 493;

Best local Similarity 48.4%; Pred. No. 9,3e-54;
Matches 182; Conservative 100; Mismatches 184; Indels 8; Gaps 3;

QY 9 MYIDQGVVWRGDWIDVVPATEAVTSRIPDQQAEDARKADAAEAPQAEWALPATER 68
 Db 5 LYIDQGVVWSSKCTVTKVSPVTCQVIGRFEATRDVDRAIDAEADAFWANDLGSVER 64
 QY 69 ASWLRKISAGIRKASELSALIVEEGRKIQQLAEVEVAFADYIDYMAFWAKRYGFIIO 128
 Db 65 SKLIYRAKELIERNAELNIIMEENGKPVKEAKEEVDGVIDQIYYAEWARKLNGEVVE 124

QY 129 SDRPGENILLKRALGVTTGILPWNFFFLIAKKMAPALLTGNLIVIKPSEFTTNNALAF 188
 Db 125 GTSSHRKIFQYKVPYGLVVALTIPWNPFAQMVAPKALAPALLTGNVIVIKSSDTPGSAEWI 184
 QY 189 AKIVDEIGLPRGVNVLGRGCTVGQELAGNPKVAMVSMTGWSA-ERIMATAAKNTFV 248
 Db 185 VPKEVAGVQKVLNFTGCHGSEFGLYIVHKKVNLITMGSTATGQPIIMKASANMAK 244
 QY 249 CLEIGGKAPALVMDADLELAVKAIIVSRVINSQVNCNCAERYVYVOKGIYQDFVNRK 408
 Db 245 ILEIGGKAPVWVKDALIMENALKTLLWAKYWNAGQS-TAAEPLVYVHEIYIDTFMSRFEVL 404
 QY 309 MQAVQGNPAERNDIANGPILINAAALFPVFGKVARAVEEAPVAPGR-AVEG KUY 463
 Db 305 SPKIALGKFNAP-METLNGKALQAT-PIVEEAKESGAKITIGSDQPSLSDGYRNY 461
 QY 364 YYPPTLLLDVROEMSIHHEETFGVPLVPVAFDTLEDAISMANDSDYGLTSSYTONLV 423
 Db 362 FFLPTLIGNAGQSKIFUREIFAPVFCAP-SSVFTMYIQLANASKYGLASYLEFTKIPNII 421
 QY 424 MKAIKGLKFGETYINRENFEAMQGFHACWRKSGIGGALGRKHGJHGYLQTVVYL 477
 Db 422 FEASFRTPFGETVYNNMPGFASQCYHTFPMTGQAGFASKYGISPYIKIKNIYV 475
 RESULT 7
 Q979W0 ID Q979W0 PRELIMINARY; PRT: 493 AA.
 AC Q979W0:
 DT 01-OCT-2001 (TrEMBLrel. 18, created)
 DT 01-OCT-2001 (TrEMBLrel. 18, last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, last annotation update)
 DE Aldehyde dehydrogenase.
 GN TVI050 OR TVG1076124.
 OS Thermoplasma volcanium.
 OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatiales;
 OC Thermoplasmataceae; Thermoplasma.
 OX NCBI_TaxID=50339;
 RN [1]
 RP SEQUENCE FROM N. A.
 RC STRAIN=GSSI / DSM 4299 / JCM 9571;
 RX MEDLINE=20570466; PubMed-11121011;
 RA Kawashima T., Anano N., Koike H., Makino S. I., Higuuchi S.,
 RA Kawashima T., Yamamoto K., Yamazaki M., Kanohori K., Kawamoto T.,
 RA Numoshima T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.;
 RT "Archaeal adaptation to higher temperatures revealed by genomic
 sequence of Thermoplasma volcanium.";
 RL Proc. Natl. Acad. Sci. U. S. A. 97:14757-14762(2000).
 DR EMBL: AP000995; BAB60192.1; .
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldehyd_1.
 DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 493 AA; 54801 MW; EDE7GE8HB292154 CRC64;

Query Match 36.9%; Score 902; DB 17; Length 493;

Best local Similarity 38.4%; Pred. No. 1.1e-54;
Matches 182; Conservative 96; Mismatches 188; Indels 8; Gaps 2;

QY 9 MYIDQGVVWRGDWIDVVPATEAVTSRIPDQQAEDARKADAAEAPQAEWALPATER 68
 Db 5 LYIDQGVVWSSKCTVTKVSPVTCQVIGRFEATRDVDRAIDAEADAFWANDLGSVER 64
 QY 69 ASWLRKISAGIRKASELSALIVEEGRKIQQLAEVEVAFADYIDYMAFWAKRYGFIIO 128
 Db 65 SKLIYRAKELIERNAELNIIMEENGKPVKEAKEEVDGVIDQIYYAEWARKLNGEVVE 124
 QY 129 SDRPGENILLKRALGVTTGILPWNFFFLIAKKMAPALLTGNLIVIKPSEFTTNNALAF 188
 Db 125 GTTHIRKIFQYKVPYGLVVALTIPWNPFAQMVAPKALAPALLTGNLIVIKSSDTPGSAEWI 184

QY 189 AKIVDEIGLPGKGVNLVSRGKFTVSGFLACNPKVAMVSMIGSVSACSEKIMATAAKNITKV 248
 Db 185 VKKFTQAGVPGALNYITGRSGVIGDYIVRHKVAGLITMTGSTGTGORIMOKASANNKL 244
 QY 249 CLEGGKAPAIYMODADLELAKAIVDSRVINSQVNCACAEVYVOKYIGIYDQFVRLGEA 308
 Db 245 MLEGGKAPFVWWDADINNAKLTLLWAKYVNAQSCIAAERLYVHEDYDTFINKFVEL 404
 QY 309 MQAVQFGPAERNDFAMGPLNNAALEREVECKVARAVEEGARVAFGGKAVE-----GKGY 363
 Db 305 SKKIVGDFSRAD---MGPLNTIALKGTEEVVEEAKESGARTLYGSGKRPDISGPKYKNGY 361
 QY 364 YYPPTLLLDVROQMSIMHEETFGVLPVAFDTILEDATSMANDSDYGLTSSITONLVA 423
 Db 362 FLPLTIIDNVOKSKIFOEIFAPVIGARKVSDVEEYSLANDSKYGLASYLFTKDFNLI 421
 QY 424 MKAIKGLKFGETYINRENFEMQGFHAGNRKSGIGGADGKHGLHGYLOTQVVYL 477
 Db 422 FEAEERIEFEGELYINMFQGFHAGNYQYHTQFRLTGAGGSGRYGTHIEIKIKNIYV 475

RESULT 8

Q9RKF1 ID Q9RKF1 PRELIMINARY: PPT: 492 AA.
 AC Q9RKF1
 DT 01-MAY-2000 (TREMBlrel 13, Created)
 DT 01-MAY-2000 (TREMBlrel 13, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel 21, Last annotation update)
 DE Putative aldehyde dehydrogenase.
 GN SC03486 OR SC865.22.
 OS Streptomyces coelicolor.
 OC Bacteria: Firmicutes, Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Murphy L., Harris D.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Cerdano A.M., Parkhill J., Barrell R.G., Rajandream M.A.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Redenbach M., Kieser H.M., Denapite D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RL the 8 Mb Streptomyces coelicolor A3(2) chromosome";
 RL Mol. Microbiol. 21:77-96(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2) / M145;
 RA Bentley S.D., Chater K.P., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.P., James K.P., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo T., Hornsby T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell R.G., Parkhill J.,
 RA Hopwood D.A.;
 RL "Complete genome sequence of the model actinomycete Streptomyces
 coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 CC -1- SIMILARITY: BLASTS TO THE ALDEHYDE DEHYDROGENASES FAMILY
 DR EMBL: AL133236; CAB61810.1; -.
 DR HSSP: P05091; 1CW3
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldedh; 1.
 DR PRS118; P500070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.

DR PRS118; P500070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
 DR PRS118; P500070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
 KW Oxidoreductase.
 SQ SEQUENCE 492 AA: 52173 MW: 109455574036FD CPC64;
 Query Match: 35.5%, Score 894; tle 16; Length 492;
 Best Local Similarity 38.9%, Pred. No. 3,8e-53;
 Matches 184, Conservative 92, Mismatches 191, Indels 6, Gaps 3;
 QY 10 YLDQFVTVRSDWIDVNVFAEAVISKIPUGQAEDARKAIDAARAQPEWALPAIERA 69
 Db 19 FVAGEYELDDSHGTLVPINPANGOLVAEVPSSSSIVDRAVIAAQAQREWRGRSHVRA 78
 QY 70 SWLPKISAGIEPKASEISALIVRPSKIDGLAFVFAFTAIYIDYMAFAKPYRELEJUS 129
 Db 79 AVLFARVAVIAJVAHDELAPIVSVQSKPLSIAMRFTESAACFFHIFALISKYRAVGSMMAS 138
 QY 130 DDPGENILLEXFALGVITGII PMNPPFPLTAPKMAPALITGNTIVIKPSEFTTNNATAFA 189
 Db 139 E-PGRSLGVREPIGVVAAIILPMNPPVAIFARKVAPALMAGNAVLPKPSLTPLSALALA 197
 QY 190 KIVIFIGLPGKGVNLVSRGKFTVSGFLACNPKVAMVSMIGSVSAGEKIMAAKNITKV 249
 Db 198 KIVIFIGLPGKGVNLVSRGKFTVSGFLACNPKVAMVSMIGSVSAGEKIMAAKNITKV 257
 QY 250 LELGGKAPAIYMODADLELAKAIVDSRVINSQVNCACAEVYVOKYIGIYDQFVRLGEA 308
 Db 255 LELGGKAPAIYMODADLELAKAIVDSRVINSQVNCACAEVYVOKYIGIYDQFVRLGEA 317
 QY 310 QAVQFGPAERNDFAMGPLNNAALEREVECKVARAVEEGARVAFGGKAVE-----GKGY 365
 Db 318 ASVIFLDPDFAAGS-KIAPLVAERPPFIPVQGMVIAAATAAPVPTGSGPHEGYSQWAF 375
 QY 366 PPTLLLDVROQMSIMHEETFGVLPVAFDTILEDATSMANDSDYGLTSSITONLVA 425
 Db 377 APTVITNVPPPEMDIAPREVFGPVLPTPTTFAERAVVSAANSTAYGLTVYVTPDUSPAMR 436
 QY 426 AIKGLKFGETYINRENFEMQGFHAGNRKSGIGGADGKHGLHGYLOTQVVYL 478
 Db 437 MIDALEFGVYVQNGARPEQVGFHTWKSSGLDGLGEGHGYFKYLPKTVVP 489
 RESULT 9
 Q916M5 ID Q916M5 PRELIMINARY: PPT: 483 AA.
 AC Q916M5
 DT 01-MAR-2001 (TREMBlrel 16, Created)
 DT 01-MAR-2001 (TREMBlrel 16, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel 20, Last annotation update)
 DE Succinate-semialdehyde dehydrogenase.
 GN CAH106 PAU765.
 OS Pseudomonas aeruginosa.
 OC Bacteria, Proteobacteria, gamma subdivision, Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 15692 / PA01;
 RA MEDLINE=20437337; PubMed=10984043;
 RA Steyer C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.L., Brinkman P.S.L., Hufnagle W.C., Kowalik L.I., Lagrou M.,
 RA Garber R.L., Goffry L., Tolentino P., Westbrock-Wadman S., Yuan Y.,
 RA Brody L.L., Goffry S.N., Folger K.R., Kas A., Larbig K., Lim P.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Pelletier T., Sailer M.H., Hancock R.F.W., Lory S., Olson M.V.;
 RL "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 opportunistic pathogen";
 RL Nature 406:959-964(2000).
 RL EMBL: AE004464; AAG03654.1; -.
 DR HSSP: P51977; 1BXS.
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldedh; 1.
 DR PRS118; P500070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.

```
DR PROSITE: PS00687: ALDH4HYDF_DEHYDR_Glu: UNKNOWN_1
KW Complete proteome.
SQ SEQUENCE: 483 AA: 51622 MW: 14016440cpr0750 CR64:

Query Match      36.0%; Score 879.5; DB 16; Length 483;
Best Local Similarity 39.3%; Pred. No. 3.7e-52;
Matches 184; Conservative 96; Mismatches 187; Indels 1; Gaps 1;

QY 10 YIDQFVTRGDAMIDVVNPNATPAVTSRIPDQGAEDARKAIDAERAQPEWALPAIERA 69
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 14 YVDGAWDADNGQTIKYNNPATGEIIGSVPMKGAAETRAIEAQAALPAWALTAKEPA 73
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 70 SWLRKISAGIRKASEISALIVBEGGKIQOLAEVEVAFTADYIDYMAFWAPYEGEITQS 129
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 74 NKLRRWFDLMIENGGDLAKLMTLEQKPLAEAKGEIAYASPLEWGEAKLIYGLTIPG 133
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 130 DRPENILLKRALGVTTGILPNWFFFLIARKMAPALLTGNITVIVKPSSEFTNNIAIFA 189
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 134 ROPDKRILIVIKQIPGVTAATTPNWPFSAMITTKATPALAAGCTMVLKASQIPYSALALA 193
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 190 KIVDEILPGVFNVLVGRGEIVGQELAGNPKVAMVSMITGSVSAEKLIMATAAKNLPKVC 249
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 194 ELAPRAGIPKVFESVVTGSACVCGELTSNPVPRKLTFTGSTFTGQIMAFCAQDIPKVS 253
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 250 LEIGGKAPAIYMDADILELAKVAILVSRVINSQVCNCAERYVVGKGIYDQFVNRIGRAM 309
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 254 LEIGNAPFIVFDADLDAAVEGALISKYRNNGQTCVCANRLVVGQVYDQFVNRIGRAM 313
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 310 QAVQFGNFAERNDIAMGPLNIAALSERVEQKVARAVEEGARVAFGKAVEKTYYPPIIL 369
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 314 AKLIGNCLGAG-VTTGPLLDARAKVAVEEHLADAVSKGKGVSGKPHALGGTFEPTI 372
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 370 LLVVRQEMSIMHIEFTGVPVLPVAFDTLEDAISMANDSDYGLTSSITYTNNIVAMKATKG 429
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 373 LVDVPKNALYSKDETEGFLAPVFERKDEAEVIAVMSNTEGLASVYFAPLAPVFPVAAEQ 432
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 430 LKFTETINENFAMGFGHAGWKSGIGGACGCKHGLHGLYLTQVYVL 477
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 433 LEYGMVGINTGLISNEVAPEGGIKASGIGREGSKYGIEDYLEIKYLC 480
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
Q8X950 PRELIMINARY; PRT: 497 AA.
AC Q8X950
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Probable succinate-semialdehyde dehydrogenase
DE (EC 1.2.1.16).
GN GABD1 OR R50028 OR R501851.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group.
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:GM11000;
FX MF0101NE:21691879; PubMed:11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Manqueot S.,
RA Ariat M., Billault A., Broffier P., Camus J.C., Catolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunne S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Siquier P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RL *Genome sequence of the plant pathogen Ralstonia solanacearum.*;
RL Nature 415:497-502(2002).
DR EMBL: AL646057; CAD13556.1;
DR pfam: PF00171; aldedh_1
DR PROSITE: PS00070; ALDH4HYDF_DEHYDR_GYS: UNKNOWN_1.
DR PROSITE: PS00687: ALDH4HYDF_DEHYDR_Glu: UNKNOWN_1.
KW oxidoreductase; Complete proteome.

SQ SEQUENCE: 497 AA: 52690 MW: 5741607099f0b7c1 CR64:

Query Match      45.7%; Score 872.5; DB 16; Length 497;
Best Local Similarity 48.3%; Pred. No. 1.2e-51;
Matches 182; Conservative 100; Mismatches 174; Indels 19; Gaps 4;

QY 18 WRGLAWI...-----DVVNPAIEAVISKIDEGQAEADAKKALDAARAKAGPWEALP 64
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 19 WRTOAFIACAWTDADDCGSKDVTDPATGKVGTCGTIVAMGAETRRALDAACAAARAKWRKT 78
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 65 AIEPASMLPKISAGITPFPASFTSALIVFFPKRIGQIAFVEVAFTAIYIDYMAFWAPYEGE 124
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 79 APEPAPILPKIADIMLPHQDIAPIITATGCKKTIIPATGFIAYASFTFWPAPFAKRVVG 148
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 125 EIIQSRRPGENILLFKALGVITGILLPWNPPFLIAKKMAPALLTGNITVIVKPSSEFTNN 184
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 139 DIIAPAGGGRKIVVNKEPTGIVIAALIPWNEFLAMMIKRWGHALAAGAMVVAFALEIDYS 198
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 185 AIAFAKIVIDEIQLPGVFNIVLGRGEIVQGEIAGNPKVAMVSMITGSVSAEKLIMATAAKN 244
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 199 ALAFELAAAGVPGAGLSIVTGAQGLIGELTANIVVPRKLSFTSTAVGRLIMRQVALD 298
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 245 ITKVCLELGRFAIVTMDIAPFIAPVAVIVSRVINSQVCNCAERYVVGKGIYDQFVNR 304
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 259 VKLSLELGGNAPFIVFTADLIAAVEGAMVAKRYRNAGQTCVCANRYVVGQVYDQFVAPAR 318
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 305 IGAEMAVAVQGNFAERNDIAMGPLNIAALSERVEQKVARAVEEGARVAFGKAVEKTYYP 360
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 315 LSAVVALPVGWNAEES-VAGGHLIGPAMIKVFAIVUNAVAGAPVIGKPHALSAGS 377
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 361 KGYVPTLLIVRQEMSIMHIEFTGVPVLPVAFDTLEDAISMANDSDYGLTSSITYTNN 420
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 378 -GAFFETVIVIAWPKMLVAHIEFTGFLAALLIDETEDAAVAAANUTEGLAAYEYTRDL 436
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 421 NVAMKALKGLKPGFTYINENFAMGFGHAGWKSGIGGACGCKHGLHGLYLTQVY 476
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 437 GRAMPVSEALFSGMVGVNIGLISAPAGVGVKSGDGPESKYGIDYIEIKYV 491
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
Q8X950 PRELIMINARY; PRT: 482 AA.
AC Q8X950
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Succinate-semialdehyde dehydrogenase, NADP-dependent
DE activity.
GN GABD OR R23959 OR R253522.
OS Escherichia coli 0157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae.
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:0157:H7 / EDL933 / ATCC 700927;
RX MEDLINE:21074935; PubMed:11206551;
RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick B.A.,
RA Postal G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck F.J., Davis N.W., Lim A., Bimblant A.F., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT *Genome sequence of enterohemorrhagic Escherichia coli 0157:H7.*;
RL Nature 409:529-534(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN:0157:H7 / RMD 0509952;
RX MEDLINE:21156231; PubMed:11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tabe T.,
RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Battori M., Shinaqawa H.;
```

*Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12.*

PT DNA Ref 8-11-22(2001)
 DR EMBL: AE005445; AAG57768.1
 DR EMBL: AP002562; BAB36945.1
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldedh.1
 DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN.1
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN.1
 KW Complete proteome.
 SQ SEQUENCE 482 AA; 51811 MW; 00F30470AAFAA5 CPO64;

Query Match 35.6%; Score 870.5; DR 16; Length 482;
 Best Local Similarity 39.3%; Pred. No. 1.5e-51;
 Matches 181, Conservative 92, Mismatches 187, Indels 1, Gaps 1;

QY 11 IDQFVTRQDAWIVVNPATRAVISPFGGAEIAEAPRAITAAEPAEAEWALVAIPAS 70
 DB 15 INGEWLDANNGEVIDVNFANRLKLSVFKMADHTFAAILAANFALPFWKALAKFPAN 74
 QY 71 WLKKTISAGIPEFASISALIVEGKIGUCLAPVEVAFADYLYMAEWAPYVEGEIIGSD 130
 DB 75 ILRNWNLMEHQDULAKMLTLEQKPLAEAKGEISYAASFLWPAEGBKKLYNTIPSH 144
 QY 131 PPGENTILFPALGVITGII PNPFEEFLIAPKMAPALLIGNIIVIKPSEETINNAIAFAK 190
 DB 135 QADKRLIVKQPIGVTAATTPNFPAAITRKAGPALAAGGIMVLKPASTPFSALALAE 194
 QY 191 IVDEIGLPRVNLVGRVETVQRIAGNPKPKVAMVMTGVSASGEKIMATAAKNITKVL 250
 DB 195 LAIRAGIPAGVFNVTGSAGVGNELTSPNPKLSPTGSTGELGMEQCAKIKKVS 254
 QY 251 ELGGKAPAIMDDADLELAVKAIIVDSRVINSQVNCNAERYVQKGIYDQFVNRIGPAMQ 310
 DB 255 ELGSAPEIPEVDDADLKAEGALASKAFENAGTGVANRYLVVGVVDFRPAFKLQAVS 314
 QY 311 AVQFGNPAERNDIAMGPLINAAALREVEQKVAFAVEGKAVAFGSKAVGGYYPPTLL 370
 DB 315 KLHIGGLGKGV-VTIGLIDKRAKAVKVEHIAALEKGAIVVGVGSKAHERGSENFQETIL 374
 QY 371 LDVROEMSIMHEFTFGPVLPPVAFDTLEDAISMANDSDYGTSSITYNLNVAMKATKGL 430
 DB 374 VQVPAKYSKEETFGPLAFKLEALVIAQANTEGGLAAYFAKULSKVFRVGEAL 444
 QY 431 KFGETYINENFAMQGFHAGWKSGIGGADCKHGLHYQ 471
 DB 434 EYGVINGINTGIISNEVAPFGIKASGLGKSGSKYGIETYDLE 474

RESULT 12

Q8ZMM2 PRELIMINARY: PRT: 482 AA.
 AC Q8ZMM2
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Succinate-semialdehyde dehydrogenase 1, NADP-dependent (EC 1.2.1.16).
 DE
 GN GABD OR STM2791.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STPAIN-LT2 / SCS01412 / A100760720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali T., Danta M., Du R., Hou S., Tayman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea R., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT *Complete genome sequence of *Salmonella enterica* serovar Typhimurium

PT L02.*
 EL Nature 413:852-856(2001).
 DP EMBL: AE006827; AAG21676.1
 DP InterPro: IPR002086; Aldehyde_dehydr
 DR Pfam: PF00171; aldedh.1
 DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN.1
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN.1
 KW Oxidoreductase; Complete proteome
 SQ SEQUENCE 482 AA; 51848 MW; C58E8F508D35097D CPO64;

Query Match 35.5%; Score 868.5; DB 16; Length 482;
 Best Local Similarity 39.4%; Pred. No. 2.1e-51;
 Matches 182, Conservative 91, Mismatches 188, Indels 1, Gaps 1;

QY 10 YLDGQVIVKGIAMIVVNPATRAVISPFGGAEIAEAPRAITAAEPAEAEWALVAIPAS 69
 DB 14 FIDHFWPDAQGVIVIPVSNPANPKLGNVPKMGCAETPDALNANPALPAMPALTAKERA 73
 QY 70 SWLPKTSAGIPEFASISALIVEPQKIGUCLAPVEVAFADYLYMAFWAPYVEGEIIGSD 129
 DB 74 NILPFWNLMEHQDULAKMLTLEQKPLAEAKGEISYAASFLWPAEGBKKLYNTIPSH 133
 QY 140 DHPGENILLFPKALGVITGII PNPFEEFLIAPKMAPALLIGNIIVIKPSEETINNAIAFA 189
 DB 134 HQADKRLIVKQPIGVTAATTPNFPAAITRKAGPALAAGGIMVLKPASTPFSALALAE 193
 QY 190 KIVDEIGLPRVNLVGRVETVQRIAGNPKPKVAMVMTGVSASGEKIMATAAKNITKVC 249
 DB 194 ELAFAPEIPEVDDADLKAEGALASKAFENAGTGVANRYLVVGVVDFRPAFKLQAV 253
 QY 250 LELGGKAPAIMDDADLELAVKAIIVDSRVINSQVNCNAERYVQKGIYDQFVNRIGPAM 309
 DB 254 LELGGNAPPIPEVDDADLKAEGALASKAFENAGTGVANRYLVVGVVDFRPAFKLQAV 313
 QY 310 QVQFGNPAERNDIAMGPLINAAALREVEQKVAFAVEGKAVAFGSKAVGGYYPPTLL 369
 DB 314 NKLAVDGGLOA-DVAILDIDKRAKAVKVEHIAALEKGAIVVGVGSKAHERGSENFQETI 372
 QY 370 LDVROEMSIMHEFTFGPVLPPVAFDTLEDAISMANDSDYGTSSITYNLNVAMKATKGL 429
 DB 373 LADVPDPAKYSKEETFGPLAFKLEALVIAQANTEGGLAAYFAKULSKVFRVGEAL 432
 QY 440 KFGETYINENFAMQGFHAGWKSGIGGADCKHGLHYQ 471
 DB 433 LEYGVINGINTGIISNEVAPFGIKASGLGKSGSKYGIETYDLE 474

RESULT 13

Q8ZAF8 PRELIMINARY: PRT: 482 AA.
 AC Q8ZAF8
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Succinate-semialdehyde dehydrogenase (EC 1.2.1.16).
 DE
 GN SLI2911.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae.
 OC Salmonella.
 OX NCBI_TaxID=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.P., Pickard D., Wain J.,
 RA Churher P., Mungall K.L., Bentley S.D., Holden M.T.P., Sebaihia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies P.M., Dowd L., White N., Farrar J.,
 RA Patwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Saoria P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrrell B.G.;
 RT *Complete genome sequence of a multiple drug resistant *Salmonella*

```

RT enterica serovar Typhi CT18.":
RL Nature 413:848-852(2001).
DR EMBL: AL627276; CAD05900.1; -.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldedh; 1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN_1
KW Oxidoreductase; Complete proteome.
SQ SEQUENCE 482 AA; 51894 MW; FFF850C52F818DC7 CRC64;

Query Match
Best Local Similarity 35.1%; Score 858.5; DB 16; Length 482;
Matches 181; Conservative 91; Mismatches 189; Indels 1; Gaps 1;

QY 10 YIDGQFVTKWGUA--WIDVNNPATEAVISKIPUGDAEAPKAIADAAFAQPEWPAIPAIPA 69
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 14 FIDGWRDARGSDVIPSVPANGKPIGNVPKMAETETRDINAANPALPAWPAITAKPA 73
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 70 SWLRKISAGIRERASEISALIVEGGKIQOLAEVEVAFTADYIDYMAEWARRYEGRIQS 129
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 74 NTLRWFLMMEHODDLARLMTLEQKPLAEAKGEISYAAASELEWFAECKRIYGDITPG 133
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 130 DRPGENILFKRALGVTTGILPNWPFPEFLIARKMAPALLIGNTIVIKPSEFTTNNAL 189
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 134 HOTDKRLVVKQPIGVTAATIPWNPSSAMITRKAGPALAAGCTMWLKPASQTPFSAL 193
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 190 KIVRIGIPGVFNVLVGGTETGQELAGNPKVAMVSMTGSVSACEKIMATAAKNITKVC 249
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 194 ELAQRAQILAGVFNWVTGSGAGTIGGELTSNPIVPKLSFTGSTEIGPQLMECAKDKK 253
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 250 LELEGKAPAIYMDADILELAVKAIYDSRVINSQVNCARVYVQKGIYDQFVNRLGEAM 309
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 254 LELEGNAPFIVFDADILKAVEGALASKFPNAGQTCVCAPIYVQGVYDRFAEKLQAV 313
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 310 QAVQFGNPAERNDIAMGPLINAAALERVEOKVARAVEEGARVAFGKAVGGYVPPPT 369
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 314 NKLAVGDLQA-DVAIGPLDERAVAKVQEHADALEKGRVITGEAHLGNFQPTI 372
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 370 LLDVQKMSIMHETFGPVIYVAVFDLTLEDAISMANDSDYGLTSSTIYTONINVAMK 429
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 373 LADVPDAKAKVETFGPLFRFSDADVIRQANUTEFLAAVYFAYRDLRSRVFVGEA 432
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 430 LKFGETYINRENFAMQGFHAGWRKSGIGGADCKHGLHYLO 471
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 433 LEYGVIGINTGIIISNVAPRGGIKASGIGRGGSKYGIENYLE 474
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
Q9K7P5
ID Q9K7P5 PRELIMINARY; PRT; 475 AA.
AC Q9K7P5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 18, last annotation update)
DE Succinate-semialdehyde dehydrogenase.
GN CABD OR BH3316.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058112;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki P., Masui N.,
RA Fuji F., Hiramata C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL: AP001518; BA807035.1; -.
DR HSSP: P51977; IBXS.
DR InterPro: IPR002086; Aldehyde_dehydr.

```

```

DR Pfam: PF00171; aldedh; 1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 475 AA; 51447 MW; 9A1A2HEHFFH117 CRC64;

Query Match
Best Local Similarity 39.1%; Score 845.5; DB 16; Length 475;
Matches 184; Conservative 84; Mismatches 195; Indels 7; Gaps 4;

QY 9 MYIDGQFVTKWGUA--WIDVNNPATEAVISKIPUGDAEAPKAIADAAFAQPEWPAIPA 66
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 6 LYIDG---TWGDLDTDFVKNPATGEVGVGMINGKFAAAAAIEAQAFAIKWKRYIAA 62
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 67 BRASWLKRSAGIRERASEISALIVEGGKIQOLAEVEVAFTADYIDYMAEWARRYEGRI 126
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 63 EKANFLMTWHLLIAGKEFLAFMLIMPMCKPLAFAPFETYSASFIPWLAFFEPVYQPT 122
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 127 IOSDRPGENILFKRALGVTTGILPNWPFPEFLIARKMAPALLIGNTIVIKPSEFTTNNAL 186
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 124 VFAEAENKPIHVAKQVWVVAALIPWNPFAAMIYPMKALALAACTFVCKPAFLPIYAV 182
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 187 AFAKIVDFIGIPGVFNVLVGGTETGQELAGNPKVAMVSMTGSVSACEKIMATAAKNIT 246
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 193 KIVELTEPAFPKGVVWLVVGGAKFTGPEMSEPHVKITFTSTAVAGKIMKQSAARM 242
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 247 KVCLEGGKAPAIYMDADILELAVKAIYDSRVINSQVNCARVYVQKGIYDQFVNRL 406
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 243 NESTFLGSGAPMILTEADILKAVAVVVAQVYVNAQVQVNVNLYVAVTEIVHFKPPT 402
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 307 EAMQAVQFGNPAERNDIAMGPLINAAALERVEOKVARAVEEGARVAFGKAVGGYV 465
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 403 SKVVQVLEKMSLEGG-VHIGELTEKKYKAKKAIYDIAVAKGAVVIGGKQGRNINISY 461
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 366 PPTLLDVRQKMSIMHETFGPVIYVAVFDLTLEDAISMANDSDYGLTSSTIYTONINVAMK 425
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 362 LPFLLTVHLEMLVMQLEETEPVAFGLALAEVLEKANGIDYGLAAVLEETRYAPLEP 421
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 426 AIKGLKFGETYINRENFAMQGFHAGWRKSGIGGADCKHGLHYLOTVV 475
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 422 LSPALFPGVFWNVCAPSTIAQADPFGMKPSGIGRGGSKYGIENYLETREV 471
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
Q92TE2
ID Q92TE2 PRELIMINARY; PRT; 484 AA.
AC Q92TE2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, last annotation update)
DE Probable succinate-semialdehyde dehydrogenase (EC 1.2.1.16).
GN CABD1 OR K00014 OR SMC02780.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OX Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN 1021;
RX MEDLINE=21396507; PubMed=11481440;
RA Capella D., Barloy-Hubler F., Gouzy J., Boite G., Ampe F., Barut J.,
RA Godard P., Becker A., Bontuy M., Gadiou F., Dreyer S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss P., Lelaire V., Masny D.,
RA Pohl T., Portetelle D., Puchot A., Porneche B., Ransperger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galliard F.;
RA "Analysis of the chromosome sequence of the legume symbiont
RA Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL: AL591782; CAC41401.1; -.
DR InterPro: IPR002086; Aldehyde_dehydr.
DR Pfam: PF00171; aldedh; 1.
DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; UNKNOWN_1.
KW Oxidoreductase; Complete proteome.

```

```

SQ  SEQUENCE  484 AA: 51866 MW: AAC2988A/205D498 CRC64;

Query Match      34.4%; Score 840.5; DB 16; Length 484;
Best Local Similarity 37.4%; Pref No. 1 7a-4g;
Matches 174; Conservative 97; Mismatches 186; Indels 9; Gaps 2;

QY  20  GUAW-----LDVVNFATEAVISKIPDGQAEIDAKAIDAAAEKAQPEWEALPAIPRASW 71
    | : | | | | | : | | | | | | | | | | | | | | | | | | | | | |
Db  16  GETWIFADPKNAIEVNNPATGEIIGRPVKLCAADTPTAIEAARVQKWEAARTAKERSAV 75
    | : | | | | | : | | | | | | | | | | | | | | | | | | | | | |

QY  72  LKKISAGIPERASEISALIVFERGKIQGLAEVFAFTADYTDYMAEWAPEYEGEIIQSDR 131
    | : | | | | | : | | | | | | | | | | | | | | | | | | | | | |
Db  76  LRRWFELMIENKDDILTMEOGKPLAEATGEIVYGASFTEWFAEEARRVYGDLPVGHQ 135
    | : | | | | | : | | | | | | | | | | | | | | | | | | | | | |

QY  132 PGNLILPKPALGVTTGILPNFPFFLIAPKMAPAILTGNTIVIKPSEFTTNNAIAPAKI 191
    | : | | | | | : | | | | | | | | | | | | | | | | | | | | | |
Db  136 KDKRILVMKQPIGVVAAIIPWNEFNAMITRKAGPALAAGCAMVLPKAAQTFESAIAIAVL 195
    | : | | | | | : | | | | | | | | | | | | | | | | | | | | | |

QY  192 VDEIGLPRGVFNLVLRGETVGOELAGNPKVAMVSMITGSVSAGEKIMATAAKNITKVCLE 251
    | : | | | | | : | | | | | | | | | | | | | | | | | | | | | |
Db  196 AEPACMPKCLFSVITGSAPETGAEMTSNPTVPKLTFTGSTEVGAELYPQSAATIKKLGLE 255
    | : | | | | | : | | | | | | | | | | | | | | | | | | | | | |

QY  252 LGGKAPAIVMDDADLELAVKALIVDSKVINSGVNCNAERYVYVQGIYDQFVNRLGEAMQA 311
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  256 LAGNAPFTIVFTATFLAAVESAIIAKEFNNGGUTGVCAKRIYVQGVYFAFSDKLQAQAVAK 315
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY  312 VQFGNPAERNDIAMGPLINAAALREVEOKVAKAVEEGARVAFGKAVEGKGYYPPTLL 371
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  316 LKTGNGME-DGVILGLIDQPALKKVEHVADALAKGARVVGGRRHSLGGTFYEATVLA 374
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY  372 DVRQEMSIHRETFGPIVLPVAFDTLEDAISMANDSDYGLTSSIYTONLNVAMKAIKGLK 431
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  375 DVTQAMAVARETEFGVPAPLFRFDESDVIAQANDTEFGLASYFYAKDLARVFRVAEALE 434
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY  432 EGETYINPENFAMQGFHAGWPKSGIGGADGKHGLHGYLOTQVYYL 477
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  435 YGVVGVNTGLISTAEAPFGVKLSGLGREGSKYGLIEEFMEIKYYVCL 480
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: June 24, 2003, 10:17:13
Job time: 40.6606 secs



GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: June 24, 2003, 10:02:25 ; Search time 40.2175 Seconds
(without alignments)
1656 hit Million cell updates/sec

Title: US-09-830-751-8
Perfect score: 2673
Sequence: 1 MTNPPSAQIKPGEVGLK... HYQCTKTLVSYSLKIKGLF 512

Scoring table: RUSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: A_Geneseq101002.4			
1:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*		
2:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*		
3:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*		
4:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*		
5:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*		
6:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*		
7:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*		
8:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*		
9:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*		
10:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*		
11:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*		
12:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*		
13:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*		
14:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*		
15:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*		
16:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*		
17:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*		
18:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*		
19:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*		
20:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*		
21:	/SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*		
22:	/SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*		
23:	/SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length	Description
1	2673	100.0	512 22 AAB74926
2	1714.5	64.1	506 22 AAG92818
3	1714.5	64.1	506 22 AAB79351
4	1682	62.9	506 22 AAB49208
5	1505.5	56.3	1302 22 ABG25490
6	1505.5	56.3	1393 22 ABG29625
7	1186.5	44.4	496 22 AAU34957
8	1090.5	40.8	508 18 AAP89793
9	1080	40.4	510 23 AAP39696
10	1074	40.2	371 22 ABG27689

11	952	35.7	509 22 AAB74924
12	949	35.5	509 22 AAB60140
13	946	35.0	517 23 AAM49516
14	936	35.0	517 23 AAM49517
15	935	35.0	495 22 AAB74923
16	924.5	34.6	497 7 AAF50452
17	919	34.4	501 21 AAG36239
18	919	34.4	501 21 AAY57413
19	918	34.3	501 21 AAY57412
20	912	34.1	477 22 AAM04880
21	905	33.9	512 23 AAF1842
22	905	33.9	512 23 AAO17364
23	905	33.9	529 22 ABG06577
24	892.5	33.4	534 21 AAY57414
25	886	33.1	199 22 AAU87144
26	873.5	32.7	654 22 AFB62256
27	873	32.7	496 22 AAU34148
28	873	32.7	496 22 AAU37064
29	870	32.5	495 16 AAR71803
30	867.5	32.5	507 23 AAM48909
31	852.5	31.9	521 15 AAR63672
32	847	31.7	485 23 AAO15196
33	847	31.7	485 23 AAU97141
34	847	31.7	508 21 AAE02082
35	847	31.7	508 23 AAE16744
36	845	31.6	496 22 AAG81968
37	845	31.6	505 17 AAR97731
38	845	31.6	506 23 AHP39538
39	844.5	31.6	517 23 AHP41923
40	843	31.5	450 22 AAU36509
41	841.5	31.5	923 23 ABB08241
42	838	31.4	487 22 AAM50256
43	837.5	31.4	496 16 AAF71891
44	834	31.2	424 21 AAG36240
45	832	31.1	496 22 AAM80032

ALIGNMENTS

RESULT 1
AAB74926
TP AAB74926 standard; Protein; 512 AA.
XX AAB74926;
XX AC
XX 26-JUN-2001 (first entry)
XX
XX E. coli aldehyde dehydrogenase-11RP protein sequence SEQ ID NO:8.
XX
XX Aldehyde dehydrogenase, glycerol dehydrogenase, 3 HP, glycerol;
XX feedstock, 3 hydroxypropionic acid, genetic engineering; glucose;
XX bacterial host, absorbable prosthetic device; surgical suture;
XX beta-lactam; acrylic acid; trifluoromethylated alcohol; diol;
XX polyhydroxyalkonate; copolymer; lactic acid.
XX Escherichia coli.
XX
XX WO200116346-A1.
XX
XX PD 08-MAR-2001.
XX
XX PF 30-AUG-2000; 2000WO-US23878
XX
XX PF 30-AUG-1999; 99QS-0151449.
XX
XX (WISC) WISCONSIN ALUMNI RES FOUND.
XX
XX PA
XX
XX FI
XX Suthers PF, Cameron DC,
XX
XX WPI; 2001-315988/33.
XX
XX DR N-PSDB; AAF82084.
XX

PT 3-Hydroxypropionic acid preparation, for use e.g. as monomer, by
 PT fermenting recombinant microorganisms expressing genes for suitable
 PT enzymes in the presence of glycerol or glucose -

XX Claim 5; Page 46-48; 63pp; English.

CC The present invention describes a method for the production of
 CC 3-hydroxypropionic acid (3-HP). The method comprises fermenting a
 CC recombinant microorganism in the presence of a source of glycerol (I)
 CC or glucose, where the microorganism: (i) expresses genes for non-native
 CC enzymes which catalyse the production of (3-HP) from (I); (ii) carries
 CC genetic constructions for the expression of a glycerol dehydratase
 CC (GDH) and aldehyde dehydrogenase (ADH) capable of catalysing the
 CC production of (3-HP) from (I); or (iii) carries a genetic construct
 CC which expresses the dhap gene from *Klebsiella pneumoniae* and a gene for
 CC an ADH capable of catalysing the production of (3-HP) from (I). 3-HP is
 CC a monomer, and is useful e.g. in the production of absorbable prosthetic
 CC devices and surgical sutures or for incorporation into beta-lactams,
 CC production of acrylic acid or formation of trifluoromethylated alcohols
 CC or diols, polyhydroxyalkonates and copolymers with lactic acid
 CC incorporation of genes encoding two enzymes makes the host organisms
 CC able to produce (3-HP) from (I). The biotechnological method of
 CC preparing (3-HP) is potentially cheaper than chemical synthesis. The
 CC present sequence represents the *E. coli* aldehyde dehydrogenase aldH,
 CC which is used in the exemplification of the present invention.

XX Sequence 512 AA;

Query Match 100.0%; Score 2673; DB 22; Length 512;

Best Local Similarity 100.0%; Pred. No. 1.3e-264;

Matches 512; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 1 MTNPPSAQIKPGEYGPPLKARYDNF1GGSEWVAFADGGEYVNTPTVTGQLLCEVASSG 60
 DB 1 MTNPPSAQIKPGEYGPPLKARYDNF1GGSEWVAFADGGEYVNTPTVTGQLLCEVASSG 60
 QY 61 KRDLIDALDAHKVKKWAHTSVQDRAAILFKIADRMENONLELLATAETWNGKPIRTS 120
 DB 61 KRDLIDALDAHKVKKWAHTSVQDRAAILFKIADRMENONLELLATAETWNGKPIRTS 120
 QY 121 AADVPLAIDHFRFASCIQAEGGISEVDSETVAYHFHEPLGVVGQIIPWNPFLLMASWK 180
 DB 121 AADVPLAIDHFRFASCIQAEGGISEVDSETVAYHFHEPLGVVGQIIPWNPFLLMASWK 180
 QY 181 MAPALAGNCVWLKPARITPLSVLLMEIVGDLIPGVVNVNAGGVIGEYLATSKRTA 240
 DB 181 MAPALAGNCVWLKPARITPLSVLLMEIVGDLIPGVVNVNAGGVIGEYLATSKRTA 240
 QY 241 KVAFTGSTEVCQIMQVATONIIIPVTILEGKSPNIVFAIVMDEEDAFDFALEGFALFA 300
 DB 241 KVAFTGSTEVCQIMQVATONIIIPVTILEGKSPNIVFAIVMDEEDAFDFALEGFALFA 300
 QY 301 FNOEVCVTCPSRALVQESIVERPMERAIIPVVESTIPSCNPLDSVTOMGACVSHGLETLIN 360
 DB 301 FNOEVCVTCPSRALVQESIVERPMERAIIPVVESTIPSCNPLDSVTOMGACVSHGLETLIN 360
 QY 361 YIDIGKKEGFAWLTGSPKKILEGELKGYLEPTILFPGNNMKVFQEEIFGVLAATTEK 420
 DB 361 YIDIGKKEGFAWLTGSPKKILEGELKGYLEPTILFPGNNMKVFQEEIFGVLAATTEK 420
 QY 421 TMEALELANDTQYGLGAGWSRNGNLATKMGRTQAGRVWNTCYHAYPAHAAFGGYKQS 480
 DB 421 TMEALELANDTQYGLGAGWSRNGNLATKMGRTQAGRVWNTCYHAYPAHAAFGGYKQS 480
 QY 481 GIGRETHKMKI FHYQOTKTIIIVSYSDKPLGLF 512
 DB 481 GIGRETHKMKI FHYQOTKTIIIVSYSDKPLGLF 512

RESULT 2

AAG92818

ID AAG92818 standard; Protein: 506 AA.

XX

AC

AAG92818;

XX

26-SEP-2001 (first entry)

XX

C glutamicum protein fragment SEQ ID NO: 6572.

XX

Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 organic acid synthesis.

XX

Corynebacterium glutamicum.

XX

EP1108790-A2.

XX

20-JUN-2001.

XX

18-DEC-2000; 2000EP-0127688.

XX

16-DEC-1999; 99JP-0377484.

XX

07-APR-2000; 2000JP-0159162.

XX

03-AUG-2000; 2000JP-0280988.

XX

(KYOWA) KYOWA HAKKO KOGYO KK.

XX

Nakagawa S, Mizouchi H, Ando S, Hayashi M, Ichiai K, Yokoi H;
 Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX

WPI: 2001 376931/40.

XX

N-PSDB: AHH8037.

XX

Novel polynucleotides derived from coryneform bacteria, for identifying
 mutation point of a gene, measuring expression of a gene, analysing
 expression profile or pattern of a gene and identifying homologous gene

XX

Claim 17; SEQ ID NO: 6572; 246pp + Sequence Listing; English.

XX

The present invention provides a number of nucleotide and protein
 sequences from the coryneform bacterium *Corynebacterium glutamicum*. These
 are useful for identifying the mutation point of a gene derived from a
 mutant of coryneform bacterium, measuring expression amount and
 analysing the expression profile or expression pattern of a gene derived
 from coryneform bacterium, and identifying a homologue of a gene derived
 from coryneform bacterium. Coryneform bacteria are useful for producing
 amino acids, nucleic acids, vitamins, saccharides and organic acids,
 particularly L-lysine. The present sequence is a protein described
 in the exemplification of the invention.

XX

Note: The sequence data for this patent did not form part of the printed
 specification, but was obtained in electronic format directly from the
 European Patent office.

XX

Sequence 506 AA;

XX

Query Match 64.1%; Score 1714.5; DB 22; Length 506;

XX

Best Local Similarity 62.5%; Pred. No. 2.2e-166;

XX

Matches 313; Conservative 82; Mismatches 105; Indels 1; Gaps 13

QY

12 PGEGYGPPLKARYDNF1GGSEWVAFADGGEYVNTPTVTGQLLCEVASSCKRDIIDALAA 71

DB

7 PGEGSVNVYKRYENYIGCKWVYFVGVYENYENYENYENYENYENYENYENYENYENYENY 66

QY

72 HKYDKWARIISVCGRAAIIIFKIAKMEONLELLATAETWNGKPIRTSAAIVPLATIDHF 131

DB

67 HAAADAWKRTSVARALILRIADRMEEHEERIAVETWENKAVRETLAAGIPLATIDHF 126

QY

132 RYFASCTRAEGGISEVDSETVAYHFHEPLGVVGQIIPWNPFLLMASWKMAFALAAAGV 191

DB

127 RYFAGAIQAEDRSSQIDHNTVAYHFHEPLGVVGQIIPWNPFLLMATWKLAFALAAAGNA 186

QY

192 VLKPAKLEPLSVLLMEIVGDLIPGVVNVNAGGVIGEYLATSKRTAKVAFIGSTVG 251

DB

187 VMKPAQIIPASIIYLLINIIGDLIPGVVNVNAGGVIGEYLATSKRTAKVAFIGSTVG 246

QY

252 QQIMQVATONIIIPVTILEGKSPNIVFAIVMDEEDAFDFALEGFALFAAGVGFVTPS 311


```

DB 247 KLINRAASUKIIEVILELGRSPSEIFSIWISQWAFASAKVSGSEFAMFALNNGEVCTGYS 309
QY 312 RALVQSEIYEPFMEPAIRVESIESNPLLSVLWMAJAVSHGULETILNYIDIGKKEGAD 371
DB 307 RALVHESIADPEFELGVKRVQNKILGNPLDTEITMNGASQEQMDKISSYKIGPPEGAG 366
QY 372 VLTGGGRKLEGLKLGYYLEPTILFGUNNMHVEEIEFGPLAVTTEKIMEALELAND 431
DB 367 TLGGKVNKVDG-MENGYIETPTVFGTNMWFIFPRITFGVLSVATFSDFDEAIRLAND 425
QY 432 TQYGLGAGVMSRNGNLAYKMRGTAQGRVWVWVYHAYPAHAAPGYYKSGIGRETHKML 491
DB 426 TNYGLGAGVMSRQNTIYRAGRAIOAGRVWVWVYHAYPAHAAPGYYKSGIGRETHKML 485
QY 492 ERYQOTKCLLVSDKPLGLF 512
DB 486 NRYQOTKCLLVSDPNPTGLF 506

RESULT 3
AAB79351
ID AAB79351 standard; Protein; 506 AA.
AC AAB79351;
XX DT 30-APR-2001 (first entry)
XX DE Corynebacterium glutamicum SMP protein sequence. SEQ ID NO 218.
XX KW Corynebacterium glutamicum, carbon metabolism and energy production.
KW SMP protein, sugar metabolism and oxidative phosphorylation protein,
KW fine chemical production, organic acid, proteinogetic amino acid,
KW nonproteinogetic amino acid, purine base, pyrimidine base, nucleoside,
KW nucleotide, lipid, saturated fatty acid, unsaturated fatty acid, diol,
KW carbohydrate, aromatic compound, vitamin; cofactor; polyketide; enzyme;
KW diagnosis; Corynebacterium diphtheriae; evolutionary study.
XX OS Corynebacterium glutamicum.
XX PN W0200100844 A2.
XX PD 04-JAN-2001.
XX PF 23-JUN-2000; 2000W0-1500943.
XX PR 25-JUN-1999; 9905-0141031.
PR 08-JUL-1999; 9905-1031412.
PR 08-JUL-1999; 9905-1031413.
PR 08-JUL-1999; 9905-1031414.
PR 08-JUL-1999; 9905-1031420.
PR 08-JUL-1999; 9905-1031424.
PR 08-JUL-1999; 9905-1031428.
PR 08-JUL-1999; 9905-1031431.
PR 08-JUL-1999; 9905-1031433.
PR 08-JUL-1999; 9905-1031434.
PR 08-JUL-1999; 9905-1031510.
PR 08-JUL-1999; 9905-1031562.
PR 09-JUL-1999; 9905-1031634.
PR 09-JUL-1999; 9905-1032180.
PR 09-JUL-1999; 9905-1042227.
PR 09-JUL-1999; 9905-1042230.
PR 09-JUL-1999; 9905-0143208.
PR 14-JUL-1999; 9905-1032924.
PR 14-JUL-1999; 9905-1032973.
PR 14-JUL-1999; 9905-1033005.
PR 27-AUG-1999; 9905-1040765.
PR 31-AUG-1999; 9905-0151572.
PR 03-SEP-1999; 9905-1042076.
PR 04-SEP-1999; 9905-1042079.
PR 04-SEP-1999; 9905-1042087.
PR 03-SEP-1999; 9905-1042088.

```

```

PR 04-SEP-1999; 9905-1042095.
PR 03-SEP-1999; 9905-1042123.
PR 03-SEP-1999; 9905-1042125.
XX (BAD) BASF AG.
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX WPI; 2001-061975/07.
XX N-PSDB; AAF71468.
XX New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
XX metabolism and oxidative phosphorylation protein for production or
XX modulation of production of fine chemicals e.g. amino acids,
XX carbohydrates or enzymes -
XX Claim 20; Page 450-452; 1246pp; English.
XX AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
XX metabolism and oxidative phosphorylation (SMP) proteins given in
XX AAB79243 to AAB 79633 which are involved in carbon metabolism and
XX energy production. The C. glutamicum SMP gene can be used in vectors
XX (II) for expression in host cells and production or modulation of
XX production of fine chemicals, such as, an organic acid, a proteinogetic
XX or nonproteinogetic amino acid (preferred), a purine or pyrimidine base,
XX a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
XX acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
XX cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
XX (II) encoded by them are used for diagnosing the presence of activity of
XX Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
XX containing them are used to map genomes of organisms related to
XX C. glutamicum, identify and localise C. glutamicum sequences of interest,
XX in evolutionary studies, in determining SMP protein regions required
XX for function, in modulating SMP protein activity, in modulating the
XX metabolism of sugars, and in modulating high energy molecule production
XX in a cell (i.e. ATP, NADPH).
XX SQ Sequence 506 AA;
Query Match 64.1%; Score 1714.5; DB 22; Length 506;
Best local Similarity 62.5%; Pred. No. 2,2e-166;
Matches 313; Conservative 82; Mismatches 105; Indels 1; Gaps 1;
QY 12 PGEYGFPLKAPYDNFVIGGEWVADGYYONLTPVTGQLCEVASSGKRDIDLALDA 71
DB 7 PTEGSTVNVYKPYENYIGKRWVPPVSGYLENISPVTCEVFEVARGAATAFVETL 66
QY 72 HKYKDKKARTSVQDPAALIFPKTADMEQNLLELAFWUNCKPIPETSAADVPLAIDHF 131
DB 67 HAAADAWGKTSVAEPALILHPTADPMEEHLEIAVETWENTKAVPETLAADIFLAIDHF 126
QY 132 RYFASCTRAQEGGISEVDSETVAYHFHEPLGVVGQIIPWFPFLLMASKMAPALAAAGNCV 191
DB 127 PYFAGATPAQEDPSSQIDHNTVAYHFNEPIGVVGQIIPWFPFLMATWKLAPALAAAGNAI 186
QY 192 VLKPAKLPLSLVLLMEIVGDLPLPGVWVWVWVWVWVWVWVWVWVWVWVWVWVWVWV 251
DB 187 VMKFAEGTASLYLINIISGLIPEVNLVNSLNGSEATLAALSSNFRTEKTAFTGSEFWG 246
QY 252 QQIMOVATGNITPVTLELSEKSFNIVADVMEESVAFPFKALEFALFNGEGVCTGYS 311
DB 247 KLINFAASUKILIVILELGRSPSEIFSIWISQWAFASAKVSGSEFAMFALNNGEVCTGYS 306
QY 312 RALVQSEIYEPFMEPAIRVESIESNPLLSVLWMAJAVSHGULETILNYIDIGKKEGAD 371
DB 307 RALVHESIADPEFELGVKRVQNKILGNPLDTEITMNGASQEQMDKISSYKIGPPEGAG 366
QY 372 VLTGGGRKLEGLKLGYYLEPTILFGUNNMHVEEIEFGPLAVTTEKIMEALELAND 431
DB 367 TLGGKVNKVDG-MENGYIETPTVFGTNMWFIFPRITFGVLSVATFSDFDEAIRLAND 425
QY 432 TQYGLGAGVMSRNGNLAYKMRGTAQGRVWVWVYHAYPAHAAPGYYKSGIGRETHKML 491

```


XX 21-MAR-2001; 2001W01050180
 PF 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 24-JUN-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX (ELIT-) ELITRA PHARM INC.
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX WPI: 2001-611495/70.
 DR N-PSDB; AAS52816.
 XX
 PT New polynucleotides for the identification and development of
 XX antibiotics, comprise sequences of antisense nucleic acids -
 PS Example 3; Seq ID No 10550; 511pp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pat_sequences.
 XX
 XX Sequence 496 AA:
 SQ
 Query Match 44.4%; Score 1186.5; DB 22; Length 496;
 Best Local Similarity 48.6%; Pred. No. 2, 9e-112;
 Matches 238; Conservative 79; Mismatches 164; Indels 9; Gaps 4;
 QY 25 YDNFIFGHVWAPANGVEYYQNIPTVTGQLLCEVASSCKKFDLALDAHKVKUKWAHTSVQ 84
 DB 14 YQLYINGEWITGSCNKKIASYNSNGEKLAEFFVDATNADVDRAVEAAQAFQTKWDVWV 73
 QY 85 DRAALFKIAURMEONLELLATARTWNGKPIRETSAADVPLAIDHFRYFASCIPAEQEG 144
 DB 74 TRSNLLIKLADLIENOEHLAMVETLNGKPKRETSTGVPSADHFRYFASVIRGEES 133
 QY 145 ISEVDSFVAYHFRPGVGVGQIIIPNFPILLMASWKMALPAAGNCVWLKPARLTPLSVL 204
 DB 134 VKFEKDTLSIVKREPCVGVGQIIIPNFPILLMGAWKLAPALAAAGNTVVIHPSSSTLSLL 193
 QY 205 LIMEIVGDLINCVNWNVNGAGGVIGHYLATSRIKAKVAFSTGVGGQIMVATQNIIP 264
 DB 194 ELFKIFQVLPKGVNIIITGRGSDSGNYMLAHGPFDKLAPTSSTGVGYTVAKAAADRIIP 253
 QY 265 VTIELGKSPNIVPADVMDDEDAFFKALRGFALFAFNGGVEGTCPSBALVQESIYEP 323
 DB 254 ATLEGGKSNIPF.....EDANWERALRGVGLGILFNGQVCCAGSRVFGSGIYDF 307
 QY 324 MERAIRRVESIRSNPLDSVTOMCAQVSHQVLEITILNYIDIGKKGADVLTGGRKRLRG 383
 DB 308 VEALKERFEQVNGFPWPKDQVEMGAQINHQLEILKYVEIGVKEGATLITGGQR-LTEN 366

QY 384 ELKDYVLEFTEILF GGNHMEVIVGELITVLAIVTHIKIMLEALILANGCYGLACVWS 442
 DB 367 GLDKGAFIAPTILLANETINIEVAQFFTEGVAIVIKFETEEFVIRIANISEYGLDAVES 426
 QY 443 KGNILAYKMGKGLGAGVWVNTNYHAYFAHAFGGYKSGSGPETHKMMLEHYGGTCLLV 502
 DB 427 GQIRVALEVAEIVPTGPMWVNYHLLAAGAH 533 AKSGSISFELHSSMLIAYTUMKNYI 496
 QY 503 SYSOKPLGLF 512
 DB 487 VTKEADGLY 496
 RESULT 8
 AAM89794
 ID AAM89794 standard; Protein: 508 AA.
 XX
 AC AAM89794;
 XX
 DT 1b-MAR-1999 (first entry)
 XX
 DE Staphylococcus aureus protein SEQ ID #5241.
 XX
 KW Computer readable medium; vaccine; *S. aureus* infection; immunodetection;
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 KW skin infection; surgical wound infection; scalded skin syndrome;
 KW toxic shock syndrome.
 XX
 OS Staphylococcus aureus.
 XX
 PN EP/86519-A2.
 XX
 PD 30-JUL-1997.
 XX
 PP 07-JAN-1997; 47EP-0100117.
 XX
 PR 05-JAN-1996; 96US-0009861.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Barash SC, Choi GH, Dillon PJ, Fannon MK, Kunsch CA;
 PI Rosen CA;
 XX
 DR WPI: 1997 374922/35.
 XX
 PT Polynucleotide(s) and proteins derived from *Staphylococcus aureus*
 PT stored on computer readable medium and used in the production of
 PT anti-*S. aureus* vaccines
 XX
 PS Claim 23; Page 3242-3244; 4271pp; English.
 XX
 CC This sequence represents a *Staphylococcus aureus* protein sequence of the
 CC invention. The DNA sequences encoding the *S. aureus* proteins are recorded
 CC on a computer readable medium, preferably selected from a floppy or hard
 CC disk, random access memory (RAM), read only memory (ROM) or CD ROM.
 CC Homology searches using the *S. aureus* DNA sequences allows putative
 CC functions to be assigned so that protein encoding or regulatory regions
 CC of commercial, therapeutic or industrial importance can be obtained.
 CC Specifically, sequences which are likely to encode antigens have been
 CC identified and these polypeptides can be used in a vaccine composition
 CC against *S. aureus* infection. The polypeptides can also be used in a kit
 CC for the immunodetection of *S. aureus* in a sample. *S. aureus* is implicated
 CC in numerous human diseases, including cellulitis, eyelid infections, food
 CC poisoning, osteomyelitis, skin and surgical wound infections, scalded
 CC skin syndrome, toxic shock syndrome, etc. Organisms transformed with the
 CC DNA sequences can be used for recombinant production of the polypeptides.
 CC The new DNA sequences (and their fragments) are useful as primers on
 CC probes for isolating homologues of any of the *S. aureus* DNA sequences
 XX contained on the computer readable medium.
 SQ Sequence 508 AA;

XX WO200175067-A2.
 PN
 XX
 PD 11-OCT-2001.
 XX
 XX 30-MAR-2001; 2001WO-US08631.
 XX
 XX 31-MAR-2000; 2000US-0540217.
 PR 24-AUG-2000; 2000US-0649167.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX Drmanac RT, Liu C, Tang YT;
 PI
 XX WPI: 2001-639362/73.
 DR N-PSDB; AAS91876.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 XX Claim 20; SEQ ID No 58048; 103pp; English.
 PS
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. Anguillo-AR330377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.
 XX
 SQ Sequence 371 AA:
 Query Match 40 28; Score 1074; DR 22; Length 471.
 Best Local Similarity 99 58; Pred No. 5, 9c-101;
 Matches 205; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 301 FNGGEVCTCPSSALVQESIVFERMERAIRRVESIRSGNPLISVTOMCAQVSHGQLETLN 360
 Db 165 FNGGEVCTCPSSALVQESIVFERMERAIRRVESIRSGNPLISVTOMCAQVSHGQLETLN 224
 QY 361 YIDICKKKGADVITCGPRKILLEGKGYVLETLPGQNNMVFQEEIFGCVLAVTTPK 420
 Db 225 YIDICKKKGADVITGPRKILLEGKGYVLETLPGQNNMVFQEEIFGCVLAVTTPK 284
 QY 421 TMEAELELANDTQYGLGAGVWSRNGNLAYKMGKRGIOAGRWNTNCHYHAYPAHAFAFGYKOS 480
 Db 285 TMEAELELANDTQYGLGAGVWSRNGNLAYKMGKRGIOAGRWNTNCHYHAYPAHAFAFGYKOS 344
 QY 481 GIGRETHKMMLLEHYQOTKCLLVSYSD 506
 Db 345 GIGRETHKMMLLEHYQOTKCLLVSYSD 370
 RESULT 11
 AAB74924
 ID AAB74924 standard; Protein: 500 AA.
 XX

AC AAB74924;
 XX 26-JUN-2001 (first entry)
 DT Human aldehyde dehydrogenase ALDH2 protein sequence SEQ ID NO:4.
 DE
 XX
 XX Aldehyde dehydrogenase; glycerol dehydratase; 4 HP; glycerol;
 KW feedstock; 3-hydroxypropionic acid; genetic engineering; glucose;
 KW bacterial host; absorbable prosthetic device; surgical suture;
 KW beta-lactam; acrylic acid; trifluoromethylated alcohol; diol;
 KW polyhydroxyalkonate; copolymer; lactic acid.
 XX
 XX Homo sapiens.
 OS
 XX WO200116346-A1.
 PN
 XX 08-MAR-2001.
 PD
 XX 30-AUG-2000; 2000WO-US24878.
 PF
 XX 30-AUG-1999; 99US-0151440.
 PR
 XX (WISC) WISCONSIN ALUMNI RES FOUND.
 PA
 XX Suthers PF, Cameron DC;
 XX WPI: 2001-315988/33.
 DR N-PSDB; AAF82082.
 XX
 XX 3-Hydroxypropionic acid preparation, for use e.g. as monomer, by
 PT fermenting recombinant microorganisms expressing genes for suitable
 PT enzymes in the presence of glycerol or glucose
 XX
 XX Claim 5; Page 32-35; 63pp; English.
 XX
 XX The present invention describes a method for the production of
 CC 3-hydroxypropionic acid (3-HP). The method comprises forming a
 CC recombinant microorganism in the presence of a source of glycerol (I)
 CC or glucose, where the microorganism: (i) expresses genes for non native
 CC enzymes which catalyse the production of (3-HP) from (I); (ii) carries
 CC genetic constructions for the expression of a glycerol dehydratase
 CC (GDH) and aldehyde dehydrogenase (ADH) capable of catalysing the
 CC production of (3 HP) from (I); or (iii) carries a genetic construct
 CC which expresses the dhap gene from Klebsiella pneumoniae and a gene for
 CC an ADH capable of catalysing the production of (3 HP) from (I). 3 HP is
 CC a monomer, and is useful e.g. in the production of absorbable prosthetic
 CC devices and surgical sutures or for incorporation into beta-lactams,
 CC production of acrylic acid or formation of trifluoromethylated alcohols
 CC or diols, polyhydroxyalkonates and copolymers with lactic acid.
 CC Incorporation of genes encoding two enzymes makes the host organisms
 CC able to produce (3-HP) from (I). The biotechnological method of
 CC preparing (3 HP) is potentially cheaper than chemical synthesis. The
 CC present sequence represents the human aldehyde dehydrogenase ALDH2, which
 CC is used in the exemplification of the present invention.
 XX
 SQ Sequence 500 AA:
 Query Match 45 78; Score 954; DR 22; Length 500;
 Best Local Similarity 42 08; Pred. No. 2, 4c 88;
 Matches 206; Conservative 84; Mismatches 176; Indels 24; Gaps 10;
 QY 28 FFGGFVWVAPAGFYQNIITVTGJLCEVASSKRPDLALD- AAIKVKIKWAIISVC 84
 Db 24 FFINNEHDVSRKTEPTPTNPTSTGEVTCQVAGSKDEIVDKAREGRGAFQAGISFWRHMVAS 84
 QY 85 DRAALFKIAURMHONLELLATATFTWNGKPIRETSAAIVPLALDHFHFYFAS- IFAQEGS 144
 Db 84 HSGELINPLALDLEKPEFTYLALELLINRKYVISTLVGLDWLELTPFYA-WAAYRPRK 144
 QY 145 ISEVDSFTVYHFIEPLGVVQDIIIPNFTPLMASWMAIPALAGN-VVLIKARGLHLSV 204
 Db 144 TIPIDGDFSTYRHEPVGVSQIIPNFTPLMOAWKLGALATGNNVVMKVAFTLAL 204

Search completed: June 24, 2003, 10:13:47
Job time : 42.2175 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - protein search, using sw model

Run on: Jun 24, 2003, 10:13:55 Search time: 14 1793 seconds
(without alignments)
1062 435 Million cell updates/sec

Title: US-09-830-751-8
Perfect score: 2673
Sequence: 1 MTNPPSAQINPGEVDFPK HYDQKTHVSYNURKPLSLF 512

Scoring table: ELCSUM2
Gap 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Issued_Patents_AA:*

- 1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/1aa/6R_COMB.pep.*
- 5: /cgn2_6/ptodata/1/1aa/PTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/1aa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1080	40.4	510	4	US-09-134-001C-4541
2	857.5	32.1	521	4	US-09-221-294.2
3	847	31.7	485	4	US-09-651-941-9
4	847	31.7	485	4	US-09-955-597-9
5	847	31.7	508	4	US-09-655-270A-9
6	845	31.6	506	4	US-09-134-001C-4383
7	752	28.1	497	1	US-08-513-841-2
8	752	28.1	497	2	US-08-646-844-2
9	752	28.1	497	2	US-08-942-673-2
10	752	28.1	497	2	US-09-118-317-2
11	731.5	27.4	488	4	US-09-134-001C-4246
12	654	22.4	444	4	US-09-134-001C-4388
13	593.5	22.2	518	4	US-09-134-001C-4451
14	583	21.8	482	4	US-09-155-183-4
15	578.5	21.6	487	4	US-09-351-224E-5
16	380.5	14.2	464	4	US-09-134-001C-4701
17	287	10.0	133	1	US-08-345-611-2
18	267	10.0	133	2	US-08-794-494-2
19	177.5	6.6	551	4	US-08-952-961-2
20	106	4.0	523	4	US-08-613-964-2
21	98.5	3.7	637	1	US-08-235-838-14
22	98.5	3.7	637	2	US-08-465-473B-14
23	97.5	3.6	711	1	US-08-235-838-7
24	97.5	3.6	711	2	US-08-465-473B-7
25	97	3.6	601	1	US-08-333-358-14
26	97	3.6	601	1	US-08-463-694-14
27	97	3.6	601	1	US-08-694-501-14

ALIGNMENTS

RESULT 1

US-09-134-001C-4541

Sequence 4541, Application US/09134001C

Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC

TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC-007

CURRENT APPLICATION NUMBER: US/09/134,001C

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 4541

LENGTH: 510

TYPE: PRT

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4541

Query Match	40.4%	Score 1080;	DR 4;	Length 510;
Best Local Similarity	43.8%	Pred No 17e+108;		
Matches 215;	Conservative 102;	Mismatches 167;	Indels 10;	Gaps 5;
QY	25	YUNFGGFWAPAGPYYQNLTPVIGLLCEVASSGKRDIDLADAAHKVKDKWHTSVQ	84	
DB	27	YGLTFNPFQASTSHTITVSNPANGPELAKVAVAKKQVTKAVCAAHAFUSWSKTSKE	86	
QY	85	DRAATLFLKLAURMEQNLLELATAETWNGKPIRETSADVPDLADHKKYFASCIKACGG	144	
DB	87	EPADYLLLEISREHKTEHLATVESLQNGKPYETISTITVPCAANKPYFASVLTDEGS	146	
QY	145	ISEVSESVAVNHEEPLVWQGLTFWNEFLMAMSKMAPALAGNCAVVKFAELTFLSVL	204	
DB	147	VNELQNLMSLVNPEVGVGVVAVNNEPILLASWKLCPALAAQNTVVIQPSSTPLSLI	206	
QY	205	LMELVIGDLLPGVVNVNNGAGGVIGFYLATSKRIAYVATGCTGTEVQOQIMQVATQNIIP	264	
DB	207	ELAKTFQEVLPKGVVNVLTGKSTSGTAIPHHEGVWLFSEFSTSCVYVNAQAIAERIVT	266	
QY	265	VTELGQGSFNIVFADNMDEECATFKALEGFALFAFNGCEVTCQSEALVGFSTYERF	323	
DB	267	ITLELGRSSNTIF-----GIANLQVTSVQGLIFNGSEVTSAGSELVSSINEL	320	
QY	324	MEFAIPFVESIESNPLSVTQMSQAVSHQGLFTLIVYDQKQKPTAAVITGTFPKLLE	382	
DB	321	LPLKAEENIKVSPFPEETKMSAQTPPELLKIESYIKIAEEDUKANILTGGHP-ITD	379	

Db 366 GYYPTPTLFSVTAIPIRETPVLSVLPFBSEEAIIILANGVFLAAAVFIPQV: 425
QY 447 LAYKMRGLOAGRWVNTYHAYPAHAAGGYSKGSIGRTHKMKMLHRYQVKCL 500
Db 426 RALRPAQTIDAGNVWINSWVLPASPYKGGGSSYGSGUGAAIESFIKESI 479

RESULT 4

US-09-955-597-9
; Sequence 9, Application US/09955597
; Patent No. 6461856
; GENERAL INFORMATION:
; APPLICANT: ROUVIER, PIERRE E.
; APPLICANT: WALTERS, DANA M
; APPLICANT: WALTERS, RUSS
; TITLE OF INVENTION: Genes Encoding Picric Acid Degradation
; FILE REFERENCE: BC1022 US NA
; CURRENT APPLICATION NUMBER: US/09/955,597
; CURRENT FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/152,545
; PRIOR FILING DATE: 1999-10-03
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 9
; LENGTH: 485
; TYPE: PRT
; ORGANISM: Rhodococcus erythropolis HL PM-1
US-09-955-597-9

Query Match 31.7% Score 847; DB 4; Length 485;
Best Local Similarity 39.0%; Pred. No. 3.5e-83;
Matches 185; Conservative 87; Mismatches 194; Indels 8; Gaps 5;

QY 29 IGCEWVADPGEYCYNLTPVTSGLLCEVASSGKPDIDLAIAHKYKKAHTSVQDPAA 88
Db 12 IGQLTPTSGTATFDSINPAQSGSHLSVAEATAADVAPAVEAAARATWQMPRPAPTP 71
QY 89 ILFKIADRMFONLELLATAETWNGKPIRETSAADVPLAIDHFFPYFASCIAPAOEGGISEV 148
Db 72 LMFYAAALIEBKTELQLOSPDMGKPIPELSGLIDPIMTILEYFAGLVTKIFGRTPA 131
QY 149 DSETVAYHHEPLGVVGIIIPNPFLLMSWKMAPALAAAGNCVVKPARLTPISVLLME 208
Db 132 PGRLNYTLREPIGVVGAITPNFPAVAVCAVWKIAPALAMGNAIVKPAQIAPLVPVALRE 191
QY 209 IVGDL-LPPGVNVVNGAGGVIGYVLAISKPIAKVETGTEVGGTGMVATONILPVIL 267
Db 192 LALEAGLPLPLGVNLIPGKSVAGNAIVLHPSVGVKVFPTGSTEVSGLIGPMAALPLITASL 251
QY 268 ELGKSPNIVFADVMDDEDAFDFKALGFALFAPNGEVCTCFPSKALVQESIYERPMERA 327
Db 252 ELGKKSALVAFGDSPPKAVA---AVVFQAMYS-NUGETCTAPSKLLIVERPIYDEVVELV 306
QY 328 IRVVSISGNPLDSVTOMGAGVSHGQLEIILNVIDLCKKGAALVLLGGRRKLLBELKD 387
Db 307 QARVAAARVGPLODPTIEIGPLISAPQRESVHVSIVSVSIEEGAILISGGQSS-PTGAPEQ 365
QY 388 GYLLPPTILFQON-NMRFVQEEIEEPVLAVTTEKIMEEALANDTQYLGAGVWSRNGN 446
Db 366 GYYPTPTLFSVTAIPIRETPVLSVLPFBSEEAIIILANGVFLAAAVFIPQV: 425
QY 447 LAYKMRGLOAGRWVNTYHAYPAHAAGGYSKGSIGRTHKMKMLHRYQVKCL 500
Db 426 RALRPAQTIDAGNVWINSWVLPASPYKGGGSSYGSGUGAAIESFIKESI 479

RESULT 5

US-09-655-270A-9
; Sequence 9, Application US/09655270A
; Patent No. 6329151
; GENERAL INFORMATION:
; APPLICANT: Rouvire, Pierre E.

; TITLE OF INVENTION: High Density Sampling of Differentially Expressed Prokaryotic
; FILE REFERENCE: BC011 US NA
; CURRENT APPLICATION NUMBER: US/09/655,270A
; CURRENT FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: 60/128,792
; PRIOR FILING DATE: 1999-February-19
; PRIOR APPLICATION NUMBER: 60/152,542
; PRIOR FILING DATE: 1999-September-03
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 9
; LENGTH: 508
; TYPE: PRT
; ORGANISM: Rhodococcus erythropolis HL PM-1
US-09-655-270A-9

Query Match 31.7% Score 847; DB 4; Length 508;
Best Local Similarity 39.0%; Pred. No. 3.8e-83;
Matches 185; Conservative 87; Mismatches 194; Indels 8; Gaps 5;

QY 29 IGCEWVADPGEYCYNLTPVTSGLLCEVASSGKPDIDLAIAHKYKKAHTSVQDPAA 88
Db 35 IGDLTPTSGTATFDSINPAQSGSHLSVAEATAADVAPAVEAAARATWQMPRPAPTP 94
QY 89 ILFKIADRMFONLELLATAETWNGKPIRETSAADVPLAIDHFFPYFASCIAPAOEGGISEV 148
Db 95 LMFYAAALIEBKTELQLOSPDMGKPIPELSGLIDPIMTILEYFAGLVTKIFGRTPA 154
QY 149 DSETVAYHHEPLGVVGIIIPNPFLLMSWKMAPALAAAGNCVVKPARLTPISVLLME 208
Db 155 PGRLNYTLREPIGVVGAITPNFPAVAVCAVWKIAPALAMGNAIVKPAQIAPLVPVALRE 214
QY 209 IVGDL-LPPGVNVVNGAGGVIGYVLAISKPIAKVETGTEVGGTGMVATONILPVIL 267
Db 215 LALEAGLPLPLGVNLIPGKSVAGNAIVLHPSVGVKVFPTGSTEVSGLIGPMAALPLITASL 274
QY 268 ELGKSPNIVFADVMDDEDAFDFKALGFALFAPNGEVCTCFPSKALVQESIYERPMERA 327
Db 275 ELGKKSALVAFGDSPPKAVA---AVVFQAMYS-NUGETCTAPSKLLIVERPIYDEVVELV 329
QY 328 IRVVSISGNPLDSVTOMGAGVSHGQLEIILNVIDLCKKGAALVLLGGRRKLLBELKD 387
Db 340 QARVAAARVGPLODPTIEIGPLISAPQRESVHVSIVSVSIEEGAILISGGQSS-PTGAPEQ 388
QY 388 GYLLPPTILFQON-NMRFVQEEIEEPVLAVTTEKIMEEALANDTQYLGAGVWSRNGN 446
Db 389 GYYPTPTLFSVTAIPIRETPVLSVLPFBSEEAIIILANGVFLAAAVFIPQV: 448
QY 447 LAYKMRGLOAGRWVNTYHAYPAHAAGGYSKGSIGRTHKMKMLHRYQVKCL 500
Db 449 RALRPAQTIDAGNVWINSWVLPASPYKGGGSSYGSGUGAAIESFIKESI 502

RESULT 6

US-09-134-001C-4383
; Sequence 4384, Application US/09134001C
; Patent No. 6480370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC

; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/364,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4383
; LENGTH: 506
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4383

Query Match 31.6%, Score 845; DB 4; Length 506;

Best Local Similarity 37.8%; Pred. No. 6.4e-R3;

Matches 190; Conservative 105; Mismatches 184; Indels 24; Gaps 11;

QY 20 KIKARYDNFIFGEWVADGAYQNLTPVTGQLLCEVASSGKRDIDLDLDAAHKVKD--K 77
 Db 16 KLSNR--QYIDGWEVSSNKKTRDINPYNQETIFTVAECTKEDVERAILAARSPEDGE 73
 QY 78 WAHTSVQDRAALFKIADRMQNIJELLATAETWNGKPIRETSAADVPLAIDHFRYFASC 137
 Db 74 WSLTSEVRGKVRADVAIKENREELAKLETLDGKTLFEE--SYAUMDOIHNVMFYAG-- 131
 QY 138 IRAOFGG-----ISEVDSFVAVHFHEPLGVVCOILLPNFPLIMASWKPALAAAGNCV 191
 Db 132 LADKGGELINSPINAEKVV-----KEPVGVVQIIPWNYPIIJQASWKIAPALATGCSL 187
 QY 192 VIKPARITPELVILLIMETVGLD--LPPGVVWVNGAGVIGEYLATSKRIAKVAFTGSTEV 250
 Db 188 VMKPESEITPLTTRVFLMERVGPCKTINILVAGASGVGVMSGHEEVDLYSFTGGIET 247
 QY 251 GQOIMQVATONIIPVITLGLGKSNIVFADVMDEELAFEDKALEGALLAFAPNGHVCTCP 310
 Db 248 GKIMQKQANVHTDVALELGGKPNITTFDUA--DFELAVDQALNG---GYFHAGQVCSAG 302
 QY 311 SRALVQESIVYKEMERAIKRVESIRSGNPLDSVTOMCAOVSHQOLETILNYIDICKKEGA 370
 Db 303 SKILVHNDIKDKFKKAILIDRVSKIKLNGFTQDTFMGPVLISTAHPDKIEGYNEVAKKQGA 352
 QY 371 DVLGGRPKLLEGLKPKZYVLEPTILFG--QNNHMFVDEEIPGHVLAATVTFKTMEEALELA 429
 Db 363 TIAIGKKRPERE-DLQAGLEFPEPTIVITDCDTSKRIYQEEVFGPVVVEGFADDEEAIRLA 421
 QY 430 NDTQYGLGAGVMSRNGNLAYKMGRIQAGRWNTNCHAYPAHAAPFGYKQSGIGRTHKM 489
 Db 422 NDSYGLAGALITKIDIGKAORVANKLKGITVWINDFHPYFAQAPWGKYKQSGIGRELKGE 481
 QY 490 MIEHYQOTKLLVSYSDPKPIGIF 512
 Db 482 GLEELYVSKHLLTNTNPEVDWF 504

RESULT 7

US-08-513-841-2

Sequence 2, Application US/08513841

Patent No. 5753481

GENERAL INFORMATION:

APPLICANT: Niwa, Mineo

APPLICANT: Saito, Yoshimasa

APPLICANT: Ishii, Yoshihori

APPLICANT: Yoshida, Masaru

APPLICANT: Suzuki, Hiromi

TITLE OF INVENTION: Dehydrogenase

OBTAINED FROM: Glucobacter-1

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: Obion, Spivak, McClelland, Majer & Neustadt, P.C.

STREET: 1755 Jefferson Davis Highway, Suite 400

CITY: Arlington

STATE: Virginia

COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: MS-DOS Editor

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/513,841

FILING DATE: 01-NOV-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: UK 9304700.9

FILING DATE: 08 MAR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 241461/1993

FILING DATE: 28-SEP-1993

ATTORNEY/AGENT INFORMATION:

NAME: NORMAN F. ORLON

PRACTICE NUMBER: 24,518

REFERENCE/DOCKET NUMBER: 18-909-0 PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-413-3000

TELEFAX: 703-413-2220

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 497 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

ORIGINAL SOURCE:

ORGANISM: Glucobacter oxydans

STRAIN: T-100

FEATURE:

NAME/KEY: mat peptide

LOCATION: 1..497

IDENTIFICATION METHOD: experimentally

US-08-513-841-2

Query Match 38.1%; Score 752; DB 1; Length 497;

Best Local Similarity 37.1%; Pred. No. 8e-73;

Matches 187; Conservative 80; Mismatches 197; Indels 40; Gaps 13;

QY 7 SAQIKPEYGLPKLKARYDNFIFGEWVADGAYQNLTPVTGQLLCEVASSGKRDIDLD 66
 Db 8 SLPLKPEFG-----FIDGWRACKD--FFDRSSPAIDVIVTVKIPCTREHIDE 55
 QY 67 ALDAAHKV--KDKWAHTSVQDRAALFKIADRMQNIJELLATAETWNGKPIRETSAADV 124
 Db 56 AVAAARAFENGSGWAGLAARAAVLLKAAGLLREHDDIAYWEVLEKCKPIQKAG-- 112
 QY 125 PLADH-----FRYFASCIKAGEG--ISEVDSFVAVHFHEPLGVVCOILLPNFPLIMASW 179
 Db 113 --EIDHICIACFEMAAGAARMJHGDFTNNLGGELFGWLRKPGVGLITTNWDPMTLCE 170
 QY 180 KMALPALAAGNCVVLKPKAPITPLSVLLIMETVGLD--LPPGVVWVNGAGVIGEYLATSKR 238
 Db 171 RAPHLLASGCTIVVKPAEVTSAITLLAEHLALAGLPGKGVENVVIGTGTVGAMTEHQD 230
 QY 239 IAKVAFTGSTEVQDQ--IMQVATONIIPVITLGLGKSNIVFADVMDEELAFEDKALEHA 297
 Db 231 IDMLSPGTSGVGVKSCITHAAUSNKKKGLLELAKKNPVVFALSNLEJAA DAVAFG 286
 QY 298 LFAFNQCEVCTCTSRALVQESIVYKEMERAIKRVESIRSGNPLDSVTOMCAOVSHQOLET 357
 Db 287 -ISFNDQGLVSSSPILVFSVAKFEPVAVVTKMERLPWVHPLDPLQIPLATTFAGAKT 445
 QY 458 ILNYIDIGKKEGALVITGGPKELLEG ELKQCYLEPTILEG QNNHMFVDEEIPGHVLA 414
 Db 446 ILDYIAGKAKGAKLLGGG--GLVDGKQYIQTPLFTHVKISMSHARDEIAGVIL 409
 QY 415 AVTTKTMEEALELANIYQGLGAGVMSRNGNLAYKMGRIQAGRWNTNCHAYPAHAAPF 474
 Db 400 ASFHEIDVDEAIAIANDIYGLAASVMSKDIKALAVIKRVAKGRVWNIIMSGGPEHLE 459
 QY 475 GYKQSGIGRTHKMMLJEHYQOTK 498
 Db 460 GGEKQSGWGEAGLYGVEEYQIK 483

RESULT 8

US-08-696-834-2

Sequence 2, Application US/08696844

Patent No. 5834263

Db 460 GGFKQSGWGRFASIGVEEYTGK 483

RESULT 11

US-09-134-001C-4246
 ; Sequence 4246, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09134,001C
 ; PRIOR FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 4246
 ; LENGTH: 488
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus epidermidis
 ; US-09-134-001C-4246

Query Match 27.4%, Score 7315, DR 4, Length 488;
 Best Local Similarity 35.6%, Pred. No. 13e-70;
 Matches 175; Conservative 91; Mismatches 204; Indels 21; Gaps 9;

QY 14 EYGPPIKIKARYTNIGRWAPAGGEYQNIPTVTGELGDEVASSGKPDIALDAHKK 73
 Db 10 EYN---KMPNPTKQVINGEWDSASSETIDVINGATFVPMKIAKSPNFKAVIAADK 66
 QY 74 VKDKWAHTSVQDRAIILFKIADPMQNIELIATAETWNGKPIPETSAALVPLAIDPEFY 133
 Db 67 VLEPPHSSVEEPRLILKIVKQYCNKNDLEIATIDELGAPLSVSNVHYQMGILNHETA 126
 QY 134 FASCIQAEGSIGSEVDSSETVAYHHEPGLGVVQIIPWNPFLIMASWMAFALAGN-VVL 193
 Db 127 APALINSQFFPQGRDGLV---KELGVA:GVTPWNPFLINTSLKLAAPAGSFVL 182
 QY 194 KPAITPIISVLIIMFIVDEL-LPPGVVWVWVWVWVWVWVWVWVWVWVWVWVWVWVWV 252
 Db 183 KPEETPEFAIILAEIFDKVGVPGVNLVNGDGGSGVGNPLSEHPKVPKMSFTGSGPTGS 242
 QY 253 QMCAIQLIPIVLELQSKSNVFAVWMPDE--LAFELKALEGHALPAPNADEVVTCF 410
 Db 243 KIMEKAADFKKVSLGLGKSPYIVLDDVDVVEEAAATILKKVN-----NIGQVVIAG 295
 QY 311 SRALVQESYIEREMERATRRVSTRSGNPLDSVTQMGAAQVSHGQLETILAYIDIGKKEGA 370
 Db 296 TRVLIPESIKEDYLTAVKEAFSKVGVGPREEDVAVSPILSKKQFTVQDYILDKGINEGA 355
 QY 371 DVLIGSKKLLSEELKQGYLEPTILFG-QNNMKVYQEEIPGPVLAVTTTEKTMEAELELA 429
 Db 356 ELFYGGPGPK-PEG-LDKGYFAPRTIFINVDNHMTIAQEEIFGPVMSVITYNNLDEATEIA 413
 QY 430 NDTQYGLGAGVSNRPNLAYKMGPGITQAGPVTNCTYHAYPAHAAGFGYKQSGIGPETHKM 489
 Db 414 NDTKYGLAGYVIGKUKOTLKHVAKSTIEASTIFINPAGKKP-DLFFGYKSGIGRWGDY 472
 QY 490 MLEHYQCTKCL 500
 Db 473 GTEEFLEVKSI 483

RESULT 12

US-09-134-001C-4388
 ; Sequence 4388, Application US/09144001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09134,001C
 ; PRIOR FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 4388
 ; LENGTH: 493
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus epidermidis
 ; US-09-134-001C-4388

Query Match 22.6%, Score 684, DR 4, Length 493;
 Best Local Similarity 29.4%, Pred. No. 1e-56;
 Matches 145; Conservative 103; Mismatches 204; Indels 42; Gaps 12;

QY 28 FIGGEWVAADGEYQNIPTVIGLLCEVASSGKREDIDLDAHKKVKKWAHTSVQDRA 87
 Db 25 FINNEFESQSKETMDVINPATGEAFDTITLATEEVNDALIEKSSQAQLLEWKEVPQPTR 84
 QY 88 AILPKIADPMQNLLELATAETWNGKPIPETSAADVPILADHFPYFASCIQAEGGI-- 145
 Db 85 EHVKLLIPLEKRNDEIAGLVYKQSKTLAG-AYGEIKRSISFIDYMTSLSMSKGPVLQ 143
 QY 146 SEVDSETVAYHHEPGLGVVQIIPWNPFLIMASWMAFALAGN-VVLKARLTPLSVLL 205
 Db 144 NSIANETHQI-INKVIGVTA:GVTPWNPFLINTSLKLAAPAGSFVL 202
 QY 206 LMEIV-GQLLPWV 264
 Db 203 LAELIPASTIPAGLIGVETVETVETVETVETVETVETVETVETVETVETVETVETV 262
 QY 265 VTELDGKSPNIVPAIWMHEDHAPFKALPAPNADEVVTCFQVSHGQLETILAYIDIGKKE 323
 Db 253 VNLGSGNAPVIV-----TSNALLKAVNVIATKPINNAQVQVQCPERIPFVHEDVDF 316
 QY 324 MEPAIFEVESIRSGNPLDSVTQMGAAQVSHGQLETILAYIDIGKKEGHALVLEGGPKLEG 383
 Db 317 ENKVTSKMKSLIVGQFPFNTDYGALINLQKGLSHHPVQALAKNZAALMTCGH----- 370
 QY 394 ELK-DEYKLEPTILFGQNNMP---VGEFFIFGEVLAVTTTEKTMEAELELANITQYGLGA 438
 Db 371 ELKPHQFPYAPVI INVPKQVWV 427
 QY 419 GWSNPNLAYKMGPGITQAGPVTNCTYHAYPAHAAGFGYKQSGIGKETHKM 490
 Db 428 YIFSENLTETVMTATEPLKFEVYVANCEAEVWVGYHA-----GWRSGIGGAGDGIHG 479
 QY 491 LEHYQCTKCLVSY 504
 Db 480 FEEYNTTVSYIRY 493

RESULT 13

US-09-134-001C-4451
 ; Sequence 4451, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09134,001C
 ; PRIOR FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674

```

; SEQ ID NO 4451
; LENGTH: 518
; TYPE: PRT
; ORGANISM: St
US-09-134-001C-4

```

Query Match 22.28; Score 593.5; DB 4, Length 518;
Best Local Similarity 34.08; Pred. No. 1.5e-55;
Matches 165; Conservative 83; Mismatches 202;
Indels 35; Gaps 16;

[illegible]

```

RESULT 14
US/09-155-183-4
: Sequence 4, Application US/09155183
: Patent No. 6323011
: GENERAL INFORMATION:
: APPLICANT: Narbad, Arjan
: APPLICANT: Rhodes, Michael J.C.
: APPLICANT: Gasson, Michael J.
: APPLICANT: Walton, Nicholas J.
: TITLE OF INVENTION: PRODUCTION OF VANILLIN
: FILE REFERENCE: 20747/100
: CURRENT APPLICATION NUMBER: US/09/155,183
: CURRENT FILING DATE: 1999-05-03
: EARLIER APPLICATION NUMBER: PCT/GB97/00809
: EARLIER FILING DATE: 1997-03-24
: EARLIER APPLICATION NUMBER: GB96/06187
: EARLIER FILING DATE: 1996-03-23
: NUMBER OF SEQ ID NOS: 14
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 4
: LENGTH: 482
: TYPE: PRT
: ORGANISM: Pseudomonas fluorescens
US/09-155-183-4

```

Query Match 21.88; Score 583; DB 4; Length 482;

[illegible]

RESULTS 15

US-09-351-224E-5
: Sequence 5, Application US/09351224E

: Patent No. 6388171

PATIENT NO. 0300171
: GENERAL INFORMATION:

; APPLICANT: Duwick, Jon

APPLICANT: Maddox, Joyce

APPLICANT: Madgax, Joyce
APPLICANT: Gilliam, Jacob

APPLICANT: Folkerts, Otto

: APPLICANT: Crasta Oswald

APPLICANT: CLASIA, OSWALD,
TITLE OF INVENTION: COMPOS

TITLE OF INVENTION: COMPOS

TITLE OF INVENTION: DETOX

FILE OF INVENTION: DETOX
FILE REFERENCE: 5718-111

FILE REFERENCE: 3718-111
CURRENT APPLICATION NUMBER

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

NUMBER OF SEC ID NOS. 11

; NUMBER OF SEQ ID NOS: 11
- SOFTWARE: FASTSEQ, CYTO WIN

```

; SOFTWARE: FASTSEQ FOR WIND
;
; SEQ 10 NOV 85

```

SEC ID NO 3
1 JAN 1971

LENGTH: 48/
TYPE: 1000

TYPE: PR

ORGANISM: *Exophiala spinosa*

US-09-351 224E-5

Query Match

Best Local Similarity

Matches 171; Conservative

yy
i0 IKPGFYGFLLK1.KAR

Figure 1. Schematic representation of the experimental design. The subjects were divided into two groups: the control group (n = 10) and the experimental group (n = 10). The control group received a standard diet (SD) and the experimental group received a high-fat diet (HFD). The subjects were divided into two groups: the control group (n = 10) and the experimental group (n = 10). The control group received a standard diet (SD) and the experimental group received a high-fat diet (HFD). The subjects were divided into two groups: the control group (n = 10) and the experimental group (n = 10). The control group received a standard diet (SD) and the experimental group received a high-fat diet (HFD).

Db 3 LSPDFY-----K

QY 69 ДААHKV -- KDKWAHT;

—
—
* *
* *
—

Db 55 AASVQAVKKGPMKFTGAQAAACMLKFADLAHKNAEKLARLESIPTRPVSMITHFDIPN 114
QY 127 AIDHEFYFASCIRAGQEGGISEVDSEIVAYHHEPLGVVGQIIPWNFPLLMASWKMAFALA 186
Db 115 MVSVPFYAGWADKTAGKTFFPEDNGKPNWRY-EPMGVCAGIASWNATFLYVGVKIAFALA 173
QY 187 AGNCVVKIPARLTPLSVLLMEIVGDL-LPPGVVNVVNGAGGVIGEVLATSKRIAKVAFT 245
Db 174 AGCSFTFKASEKSPICVIGIAPLFAEAGFPVGVQFLTGA-PVTGEALASHMDIAKISFT 232
QY 246 GSTEVGQIMQYAT--QNIIPVILELGGKSPNIVFADVMDEEDAFFDKALEGFALFAPNQ 303
Db 233 RSVGGGPAVKQ-ATLKSNMKPVILELGEK-PTIVNEAPLERQS--GESAKDFSKF---- 284
QY 304 GEVCTCPSPAIWQ-ESIYEPFMEPAIPRVESIPQS-----NPLDSVTOMGAQVS 351
Db 285 GQIWWPPSP*LLVQWNIARKF-----HVVPHSDFSN*QFWLQNPFPKPTHGPFVD 336
QY 352 HQOLETILNYIDIGKKEGADVLTGGRKLLLEGELKDGYYLEPTILFG-QNNMRVFOEEIF 410
Db 337 KSYDRVLGNIDWG-KDTAQLLIGVGRK---GD--KGFALEPTIFVNPKGSKIWFEEIF 390
QY 411 GPVLAVTTTFTKEEAELELANDTQYGIAGVMSRPNCAIAYKMGEGIQAGRPWTCYHAYPA 470
Db 391 GPVLSIKTKTEERAEIANDTTYGLASVIYTKSLNRLRVSSALETGGVSINPPFIPET 450
QY 471 HAAFGYKQSGIGRETHKMMLEHYQOIKCL 500
Db 451 QTPFGMKQSGSGKELGEBGLKAYLEPKTI 480

Search completed: June 24, 2003, 10:31:44
Job time : 16.1793 secs

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

697

698

699

700

701

702

703

704

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756

757

758

759

760

761

762

763

764

765

766

767

768

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807

808

809

810

811

812

813

814

815

816

817

818

819

820

821

822

823

824

825

826

827

828

829

830

831

832

833

834

835

836

837

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868

869

870

871

872

873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919

920

921

922

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

967

968

969

970

971

972

973

974

975

976

977

978

979

980

981

982

983

984

985

986

987

988

989

990

991

992

993

994

995

996

997

998

999

1000

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on Jun 24, 2003, 10 17.21, Search time 24 74.92 Seconds
(without alignments)

Title: US-09-830-751-8
Perfect score: 2673
Sequence: 1 MTNNPPSAOIKPGFYGFPLK.....HYGQTKCLIVSYSDKPIGLF 512

Scoring table: BLOSUM62

Securing cable: BLOSUM02
Gapor 10 0 , Gapext 0 5

Searched: 417779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing.	Minimum Match 0%
	Maximum Match 100%
	Listing first 45 s

Database : Published Applications AA: *

- Two-Sided Applications: PA:
- 1: /cqn2-6/pdata/2/pubpaa/us08_NEW_PUB.pcp.*
 - 2: /cqn2-6/pdata/2/pubpaa/rct_NEW_PUB.pcp.*
 - 3: /cqn2-6/pdata/2/pubpaa/us05_NEW_PUB.pcp.*
 - 4: /cqn2-6/pdata/2/pubpaa/us07_PUBCOMB.pcp.*
 - 5: /cqn2-6/pdata/2/pubpaa/us07_NEW_PUB.pcp.*
 - 6: /cqn2-6/pdata/2/pubpaa/us07_PUBCOMB.pcp.*
 - 7: /cqn2-6/pdata/2/pubpaa/rct03_PUBCOMB.pcp.*
 - 8: /cqn2-6/pdata/2/pubpaa/us08_PUBCOMB.pcp.*
 - 9: /cqn2-6/pdata/2/pubpaa/us09_NEW_PUB.pcp.*
 - 10: /cqn2-6/pdata/2/pubpaa/us09_PUBCOMB.pcp.*
 - 11: /cqn2-6/pdata/2/pubpaa/us10_NEW_PUB.pcp.*
 - 12: /cqn2-6/pdata/2/pubpaa/us10_PUBCOMB.pcp.*
 - 13: /cqn2-6/pdata/2/pubpaa/us60_NEW_PUB.pcp.*
 - 14: /cqn2-6/pdata/2/pubpaa/us60_PUBCOMB.pcp.*

Pred No is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1714.5	64.1	505	9	US-09-738-626-6572	Sequence 6572, Appl
2	1186.5	44.4	496	10	US-09-815-242-5644	Sequence 10856, A
3	1080.5	40.8	508	1	US-08-781-986A-5241	Sequence 5241, Appl
4	946.5	35.4	492	9	US-10-268-518-4	Sequence 4, Appl
5	936	35.0	493	9	US-10-175-696-21	Sequence 21, Appl
6	936	35.0	494	10	US-09-823-901-9	Sequence 9, Appl
7	919	34.4	501	9	US-09-344-882-22	Sequence 22, Appl
8	919	34.4	501	9	US-10-293-865-22	Sequence 22, Appl
9	918	34.3	538	9	US-09-344-882-20	Sequence 20, Appl
10	918	34.3	538	9	US-10-293-865-20	Sequence 20, Appl
11	905	33.9	512	9	US-09-963-403-12	Sequence 12, Appl
12	905	33.9	512	9	US-10-268-518-2	Sequence 2, Appl
13	905	33.6	512	9	US-10-205-823-14	Sequence 14, Appl
14	892.5	33.4	534	9	US-09-344-882-24	Sequence 24, Appl
15	892.5	33.4	534	9	US-10-293-865-24	Sequence 24, Appl
16	873	32.7	496	10	US-09-815-242-5644	Sequence 5644, Appl
17	873	32.7	496	10	US-09-815-242-15657	Sequence 12657, A
18	870	32.5	495	9	US-09-847-209-11	Sequence 11, Appl
19	844.5	31.6	518	9	US-09-919-039-143	Sequence 143, Appl

ALIGNMENTS

```

1 RESULT 1
2 US 09/738,626-6572
3 7 SEP-09 09:47Z APPL NO. 09/738,626
4 PUBLICATION NO. US000219769A1
5 GENERAL INFORMATION:
6 APPLICANT: NAKAGAWA, SATOSHI
7 APPLICANT: MIZOGUCHI, HIROSHI
8 APPLICANT: ANDO, SEIKO
9 APPLICANT: HAYASHI, MIKIRO
10 APPLICANT: OCHIAI, KEIKO
11 APPLICANT: YOKOI, HARUHIKO
12 APPLICANT: TATEISHI, NAKKO
13 APPLICANT: SENOH, AKIHIRO
14 APPLICANT: IKEDA, MASATO
15 APPLICANT: OZAKI, AKIO
16 TITLE OF INVENTION: N-VEL. POLYNUCLEOTIDES
17 FILE REFERENCE: 249-125
18 CURRENT APPLICATION NUMBER: US/09/738,626
19 CURRENT FILING DATE: 2000-12-18
20 PRIOR APPLICATION NUMBER: JP 99/377484
21 PRIOR FILING DATE: 1999-12-16
22 PRIOR APPLICATION NUMBER: JP 00/159162
23 PRIOR FILING DATE: 2000-04-07
24 PRIOR APPLICATION NUMBER: JP 60/280988
25 PRIOR FILING DATE: 2000-08-03
26 NUMBER OF SEQ ID NOS: 7059
27 SOFTWARE: PatentIn ver. 3.0
28 SEQ ID NO 6572
29 LENGTH: 506
30 TYPE: PR1
31 ORGANISM: Corynebacterium glutamicum
32 SEQ ID NO 6572

```

Query Match 64.16; score 1714.5, bK 9; Length 506;

Query Match 64.16; Score 174.5; DB9; Length 506;
Best Local Similarity 62.58; Pred. No. 7.1e-152;

Best Local Similarity 92.5%, FREQ. NO. 7, 10, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 843, 844

12 PREY7FFIKIKAKHYONFI3GFWWAFAC0EY YUNLIFVTGULL'EVASS3KFFDIDLALDAA 71

[illegible]

72 HFVDERWAHTSVQDRAAHLFKIADPMENLELLATAETWQDNGFIRKTSAAIVPLAIDHF 131

```

Db      67 HAAADAWCKTSVAERALLHRIADRMEEHLEETAVAEETWENKAVRETTAAADIFLADHF 126
QY      132 RYFASCIQAEQGGISEVDSEIVAYHFHEPLGVVGOIIPWNPFLIMASWKMALAAAGNCV 191
Db      127 RYFAGAIKAEQDSQIDHNTVAYHFNEPIGVVGOIIPWNPFLIMATWKLAPALAAAGNAI 186
QY      192 VLKPARLTPLSVLLMEITVGLLPPGVVNVVNGAGGVIGYLAITSKRIAKVAFTGSTEVG 251
Db      187 VMKPAEQTPASILYLINIGLDLPEGVNLVNGLGRFAGAAALSGSNRIKIAFTGSTEVG 246
QY      252 QOIMQYATONTIIPVTELEGCKSPNIVFAVDWDEDAFDFKALEGFALFAFNQGEVCTCPS 311
Db      247 KLINRAASDKIIPVTELEGCKSPSIFESDVLSDQDAFAEKAVEGFAMFALNQGEVCTCPS 306
QY      312 BALVOESTYERFMERATRVESIRSGNPLOSVMQAGVSHQGLIETILNYITFGSKKEZAL 371
Db      307 RALVHESIADFELEGVKRVONTIKLGNPLDTETMMGAQASQEQMDKISSYLKIGPEGAG 366
QY      372 VLTGRRKLLCEGLKDRYYIRPTTIFGNNMRVFQERTFGPVLAVTTFKTMEEALELANC 431
Db      367 TLTGKVKVNDG-MENGYIETPTVPGTNDMPFPEIFGVLVSATFSDFDAIRIAND 425
QY      432 TOYGLGAGVWRNENLAKMGRGTQAGRVWNTNCYHAYPAHAAGFYKQSGIGRETHKMML 491
Db      426 TNYGLGAGVWRDQNTIYRAGRAIQAGRVWNVYHYPAHSAFGYKQSGIGRENHLMML 485
QY      492 EHYQOTKCLLYSDDKPLGLF 512
Db      486 NHYQOTKLLVSDPNPTGLF 506

```

RESULT 2

```

US-09-815-242-10550
; Sequence 10550, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10550
; LENGTH: 496
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10550

```

Query Match

44.4%; Score 1186.5; DB 10; Length 496;

```

Best Local Similarity 48.6%; Prod. No. 1,90-102;
Matches 238; Conservative 79; Mismatches 164; Indels 9; Gaps 4;
QY      25 YDNFIFGRWAPAPRCFYQNI TPVTGQI LPTFVASSCKRPDIDIAL DAHKVKIKWAIHVSQ 84
Db      14 YQLYINGFTWTGSCNKKMIASYNISNGCKLAFFVDATNAIDVRAVEAQAQAFQTKWIVDV 74
QY      85 DRAALIFKIDRMFONLELATATWNGKPIRETSAADVPFLAIDHPRFASITRAQPEK 144
Db      74 TFSNLLKIDADLEENDEHLMAMVETLNGRPLPETUSIDNPASADHPRFASVIRGERES 134
QY      145 ISEVDSEIVAYHFHEPLGVVGOIIPWNPFLIMASWKMALAAAGNCV VVLPKARLPLSVL 204
Db      134 VKFEDKDTLSIVVKEPIGVVGOIIPWNPFLIMGAKWALATALAAGNTVVHIPSSSTSLSL 194
QY      205 LLMELIVGLLPPGVVNVVNGAGGVIGYLAITSKRIAKVAFTGSTEVCQOIMQYATONTIIP 264
Db      194 ELFIPTQVLPKGVVNIITGMSQSNYMLAHPRFPKLAFTGSTEFTVYVAKAAADPLIP 254
QY      265 VTLELGCKSPNIVFAVDWDEDAFDFKALEGFALFAFNQGEVCTCPSBALVOESTYERFM 424
Db      254 ATLELGCKSANIIP EDANKERALEESVQGLIPTRQGVVCAAGHAWTVQGLIYDQF 407
QY      324 MEHALIRVESIRSGNPLOSVMQAGVSHQGLIETILNYITFGCKKCADWITGRRKILDES 484
Db      308 VEALKKPEQVNVGVFPWEKIVEMCAQINERHOLEELIKYVEIRKVECATILITGQR LEN 466
QY      384 ELKQGYYLEPTILF-GONNRKVPQEEIFGVLAVTTFKTMEEALELANOTYGLGAGVWS 442
Db      367 GLDKGAPLAPTLANGTNTMCAQVAFIFGFPVATVVKPETFEERVIRLANDISEYGLGAGVFS 426
QY      443 RGNLAYKMGRTQAGRVWNTNCYHAYPAHAAGFYKQSGIGRETHKMMLHYQOTKCLLY 502
Db      427 QDINVALKVARGVHLEHMMVNYNLQLEAFAPFGYKQSGIGRETHKSMIDJAYTOMKNYI 486
QY      503 SYSDKPLGLF 512
Db      487 VTKEEADGLY 496

```

RESULT 3

```

US-08-781-986A-5241
; Sequence 5241, Application US/08781986A
; Publication No. US20030054436A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Beuson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504

```



```

; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/211,727
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: consensus sequence
US-10-175-696-21

```

```

Query Match      35.0%; Score 936; DB 9; Length 493;
Best Local Similarity 45.2%; Pred No 5, 3a-7a;
Matches 225; Conservative 79; Mismatches 160; Indels 34; Gaps 17;

QY 32 EWAPADGEYUNLPVT-GQLCEVASSGKRDIIDLALDA---AHKVKDKWAHTSVQDRA 87
DB 1 EWDSASGKTFEVVNPAKNGEIVGRVPEATAEDVDAVKAKEAFKSGPWAQVPASERA 60

QY 88 AILEKIDRMENONLELLATAETWNGKPIRETSA-ALVPLAIUHFYFASCIKAQBG--- 143
DB 61 RILRLADLIERDELAUETLDLGPKEAEAGDTEVGRADIDEIRYAGWARKLMGERR 120

QY 144 ---GISEVDSSTVAYHEHEPLGVVGQIIPWNPFLMASMKMAPALAAAGNCVVLKPARLTP 200
DB 121 VIPSLATDGDDEELNYTRREPLGVGVVSPWNPFLLLALNKLALPALAAGNTVVLKPSQTP 180

QY 201 LSVLLIMELVCDL-----LPGVVNVVNVNAGVIGEVYATSKRIAKVAFSTGVQOLIMJ 256
DB 181 LTALLAELEAGANNLPKGVNVNVPFGAEGVQALLSHPDIDKISFTGSTEVGKLIME 240

QY 257 -YATQNIIPVTLELGKSPNIVFADVMDDEDAFEDKALEGFALFAP-NOGEVCTCPSRAL 314
DB 241 AAAAKNKKVTLLEGKSPVIVF-----DDADLKAVERIVGAFGNAGOVCIAPSRLL 294

QY 315 VOESYIYFEMERAIRRVESIR-SCNPLDSVTQM-GAUVSHQULETIL-NYIDIGKREGAD 371
DB 295 VHESYIDFVFKIKERVKKLIGDPLDSDTNIYGPLISEOQFDRVLWSVIFDKGREGAK 354

QY 372 VLTGGRKLLGELGKDYLYLEPTILFCQN-NMVFQFEPFGVPLAVTTTEKTWEAELELAN 430
DB 355 VILGGERDESEYLGCGYVYQPTLFTIVTDPMKIMKEEIPGVPLPIIKFKDLEAELELAN 414

QY 431 DTQYGLGAGVWSRN-GNLAYKMGREGIQAGRWNTN--CYHAYPAHAAPGGYKQ-SGIGRET 486
DB 415 DTEYGLAAYVTKDILAPAFVAKALFAGIVWVNDVCVHAAPFOLPFGGKQSSGIGRE- 473

QY 487 H--KMMLEHYQOTKCLIV 502
DB 474 HGGKYGLEEYETIKVTI 491

```

```

RESULT 6
US-09-823-901-9
; Sequence 9, Application US/09823901
; Patent No. US20020001807A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: DEHYDROGENASES AND USES THEREOF
; FILE REFERENCE: 10448-036001
; CURRENT APPLICATION NUMBER: US/09/823,901
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/194,920
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Artificial Sequence

```

```

; FEATURE:
; OTHER INFORMATION: Consensus Sequence
US-09-823-901-9

Query Match      45.0%; Score 946; DB 10; Length 494;
Best Local Similarity 45.2%; Pred No 5, 3a-7a;
Matches 225; Conservative 79; Mismatches 160; Indels 34; Gaps 17;

QY 32 EWAPADGEYUNLPVT-GQLCEVASSGKRDIIDLALDA---AHKVKDKWAHTSVQDRA 87
DB 1 EWDSASGKTFEVVNPAKNGEIVGRVPEATAEDVDAVKAKEAFKSGPWAQVPASERA 60

QY 88 AILEKIDRMENONLELLATAETWNGKPIRETSA-ALVPLAIUHFYFASCIKAQBG--- 144
DB 61 RILRLADLIERDELAUETLDLGPKEAEAGDTEVGRADIDEIRYAGWARKLMGERR 120

QY 144 ---GISEVDSSTVAYHEHEPLGVVGQIIPWNPFLMASMKMAPALAAAGNCVVLKPARLTP 200
DB 121 VIPSLATDGDDEELNYTRREPLGVGVVSPWNPFLLLALNKLALPALAAGNTVVLKPSQTP 180

QY 201 LSVLLIMELVCDL-----LPGVVNVVNVNAGVIGEVYATSKRIAKVAFSTGVQOLIMJ 256
DB 181 LTALLAELEAGANNLPKGVNVNVPFGAEGVQALLSHPDIDKISFTGSTEVGKLIME 240

QY 257 -YATQNIIPVTLELGKSPNIVFADVMDDEDAFEDKALEGFALFAP-NOGEVCTCPSRAL 414
DB 241 AAAAKNKKVTLLEGKSPVIVF-----DDADLKAVERIVGAFGNAGOVCIAPSRLL 294

QY 315 VOESYIYFEMERAIRRVESIR-SCNPLDSVTQM-GAUVSHQULETIL-NYIDIGKREGAD 471
DB 295 VHESYIDFVFKIKERVKKLIGDPLDSDTNIYGPLISEOQFDRVLWSVIFDKGREGAK 454

QY 372 VLTGGRKLLGELGKDYLYLEPTILFCQN-NMVFQFEPFGVPLAVTTTEKTWEAELELAN 430
DB 355 VILGGERDESEYLGCGYVYQPTLFTIVTDPMKIMKEEIPGVPLPIIKFKDLEAELELAN 414

QY 431 DTQYGLGAGVWSRN-GNLAYKMGREGIQAGRWNTN--CYHAYPAHAAPGGYKQ-SGIGRET 486
DB 415 DTEYGLAAYVTKDILAPAFVAKALFAGIVWVNDVCVHAAPFOLPFGGKQSSGIGRE- 473

QY 487 H--KMMLEHYQOTKCLIV 502
DB 474 HGGKYGLEEYETIKVTI 491

RESULT 7
US-09-344-882-22
; Sequence 22, Application US/09444882
; Patent No. US20020162137A1
; GENERAL INFORMATION:
; APPLICANT: Nikolau, Basil J
; APPLICANT: Wurtele, Eve S
; APPLICANT: Oliver, David J
; APPLICANT: Behal, Robert
; APPLICANT: Schnable, Patrick S
; APPLICANT: Ke, Jinsuan
; APPLICANT: Johnson, Jerry L
; APPLICANT: Allred, Carolyn C
; APPLICANT: Faland, Beth
; APPLICANT: Lutiger, Isabelle
; APPLICANT: Wen, Tsui-Jung
; TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
; FILE REFERENCE: 201573
; CURRENT APPLICATION NUMBER: US/09/444,882
; CURRENT FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: 03/03,990,717
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.2
; SEQ ID NO 22
; LENGTH: 501
; TYPE: PRT

```



```
; ORGANISM: Arabidopsis Thaliana
US-09-344-882-22

Query Match      34.4%  Score 919, DB 9, Length 501;
Best Local Similarity 39.4%  Pred. No. 2.1e-77;
Matches 196; Conservative 100; Mismatches 165; Indels 36; Gaps 11;

QY 28 FIGGEWAPADGEYYQNLTPVTQGLCEVASSGKRDIDLALDAHAKVKD--KWAHTSVQD 85
DB 23 FIGGQIDAASGKFTETIDPRNGEVIATAGDQKEDVDLAVNAARAFDCHGPMRTGFE 82
QY 86 RAATLFIKADMEQNLLELATAETWONGKPIRETSAADVPLAIDHFRYFASCIROGGI 145
DB 83 RAKLTKNFADLLEENIELAKLDAVDGSKIPLQKGYADIPATAGHERYNA-----GAA 135
QY 146 SEVUSETV-----AYHFEHPGLGVVGGQIIPWNPFLIMASWMAFALAAGNCVVLKPAR 197
DB 136 DRHGETLKMTPQSLFGYTLKEPIGVVGNITPWNFPSPIMFATKVPAMAAAGCTMVVKPAE 195
QY 198 LPLSVLLMEIVGDL-LPPGVWVWVNGAGGIVGEYLATSKPIAKVAFTGSTEVGQOIMQ 256
DB 196 QTSLSALFYAHLKRENGIPGCVLNVITGPGSTAGAAIASHMDVKSFTGSDVGRKIMQ 255
QY 257 -YATONIIPTVLELGGKSPNIVFADVMDEEDAFDFALEGFALFA--FNOGEVCTCPSPA 313
DB 256 AAAASNLKKVSLLEGGKSPLLIF-----NDADIDKRAAD-LALLGCFYKNGEICVASSRV 308
QY 314 LVQESIYERFERAARRVESIRSGNPLDSVTOMGAQVSHGQLETLINLYIDIGKKGADVL 373
DB 309 FVQEGIVKVEKLVKRAKDWTDGDFDSTARQGPQVDKPKQFKILSYIEHGKNEGATLL 368
QY 374 TGGRRKLLEGEIKD-GYYLEPTILFG-QNNMKVQEEIFGVLAVTIFKIMEALELAND 431
DB 369 TGGK-----AIGDKGYIQTIPADVTEDMKIYQDEIFGPNVSLMKFKTVRGKICANN 422
QY 432 TOYGLGAGWSRNGNLAYKMGREGTOAGRVWNTNCYHAYPAHAAGGYKQSGIGRETHKML 491
DB 423 TRYGLAAGILSQDILINTVSPKAGIWWNVYFSGTLLCPYGGYKMSGNCPESGMDAL 482
QY 492 EHYQOTKCLLVSYSLKP 508
DB 483 DNYLQTKSVVPLHNSP 499

RESULT 8
US-09-293-865-22
; Sequence 22, Application US/10293865
; Publication No. US2003010800A1
; GENERAL INFORMATION:
; APPLICANT: Nikolau, Basil J
; APPLICANT: Wurtele, Eve S
; APPLICANT: Oliver, David J
; APPLICANT: Behal, Robert
; APPLICANT: Schnable, Patrick S
; APPLICANT: Ke, Jinshan
; APPLICANT: Johnson, Jerry L
; APPLICANT: Allred, Carolyn C
; APPLICANT: Patland, Beth
; APPLICANT: Lutziger, Isabelle
; APPLICANT: Wen, Tsui-Jung
; TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
; TITLE OF INVENTION: Acetyl CoA Levels in Plants
; FILE REFERENCE: 217113
; CURRENT APPLICATION NUMBER: US/10293865
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/344,882
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 60/090,717
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 22
; LENGTH: 501

; TYPE: PRT

; ORGANISM: Arabidopsis Thaliana
US-10-293-865-22

Query Match      34.4%  Score 919, DB 9, Length 501;
Best Local Similarity 39.4%  Pred. No. 2.1e-77;
Matches 196; Conservative 100; Mismatches 165; Indels 36; Gaps 11;

QY 28 FIGGEWAPADGEYYQNLTPVTQGLCEVASSGKRDIDLALDAHAKVKD--KWAHTSVQD 85
DB 23 FIGGQIDAASGKFTETIDPRNGEVIATAGDQKEDVDLAVNAARAFDCHGPMRTGFE 82
QY 86 RAATLFIKADMEQNLLELATAETWONGKPIRETSAADVPLAIDHFRYFASCIROGGI 145
DB 83 RAKLTKNFADLLEENIELAKLDAVDGSKIPLQKGYADIPATAGHERYNA-----GAA 135
QY 146 SEVUSETV-----AYHFEHPGLGVVGGQIIPWNPFLIMASWMAFALAAGNCVVLKPAR 197
DB 136 DRHGETLKMTPQSLFGYTLKEPIGVVGNITPWNFPSPIMFATKVPAMAAAGCTMVVKPAE 195
QY 198 LPLSVLLMEIVGDL-LPPGVWVWVNGAGGIVGEYLATSKPIAKVAFTGSTEVGQOIMQ 256
DB 196 QTSLSALFYAHLKRENGIPGCVLNVITGPGSTAGAAIASHMDVKSFTGSDVGRKIMQ 255
QY 257 -YATONIIPTVLELGGKSPNIVFADVMDEEDAFDFALEGFALFA--FNOGEVCTCPSPA 313
DB 256 AAAASNLKKVSLLEGGKSPLLIF-----NDADIDKRAAD-LALLGCFYKNGEICVASSRV 308
QY 314 LVQESIYERFERAARRVESIRSGNPLDSVTOMGAQVSHGQLETLINLYIDIGKKGADVL 373
DB 309 FVQEGIVKVEKLVKRAKDWTDGDFDSTARQGPQVDKPKQFKILSYIEHGKNEGATLL 368
QY 374 TGGRRKLLEGEIKD-GYYLEPTILFG-QNNMKVQEEIFGVLAVTIFKIMEALELAND 431
DB 369 TGGK-----AIGDKGYIQTIPADVTEDMKIYQDEIFGPNVSLMKFKTVRGKICANN 422
QY 432 TOYGLGAGWSRNGNLAYKMGREGTOAGRVWNTNCYHAYPAHAAGGYKQSGIGRETHKML 491
DB 423 TRYGLAAGILSQDILINTVSPKAGIWWNVYFSGTLLCPYGGYKMSGNCPESGMDAL 482
QY 492 EHYQOTKCLLVSYSLKP 508
DB 483 DNYLQTKSVVPLHNSP 499

RESULT 9
US-09-344-882-20
; Sequence 20, Application US/09344882
; Patent No. US2002012137A1
; GENERAL INFORMATION:
; APPLICANT: Nikolau, Basil J
; APPLICANT: Wurtele, Eve S
; APPLICANT: Oliver, David J
; APPLICANT: Behal, Robert
; APPLICANT: Schnable, Patrick S
; APPLICANT: Ke, Jinshan
; APPLICANT: Johnson, Jerry L
; APPLICANT: Allred, Carolyn C
; APPLICANT: Patland, Beth
; APPLICANT: Lutziger, Isabelle
; APPLICANT: Wen, Tsui-Jung
; TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
; TITLE OF INVENTION: Acetyl CoA Levels in Plants
; FILE REFERENCE: 201573
; CURRENT APPLICATION NUMBER: US/09/344,882
; CURRENT FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 60/090,717
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.2
; SEQ ID NO 20
; LENGTH: 538
; TYPE: PRT
```

ORGANISM: Arabidopsis Thaliana
US-09-344-882-20

Query Match 34.3%; Score 918; DB 9; Length 538;
Best Local Similarity 41.1%; Pred. No. 2,96-77;
Matches 202; Conservative 30; Mismatches 170; Indels 30; Gaps 12;

QY 29 TGGHVAADGEYQNLTPVIGQLICRVASSGRKIDIALDAAHKVKDK--WAHITSVQUR 86
DB 62 INGNFVSASCKTPTLPPPTGEVIAHVAEEDAEINRAVKAAPTAFADESPWPKMSAYER 121
QY 87 AAILFKIADRMEOQLLELATAETWNGKPIRETSAADVPLAIDHFRVYFASCIQAQEGGIS 146
DB 122 SRVLLRPADLVKHSSELASLETWNGKPYQOSITAEIPMFARLPFYAGWADKIHGLTI 181
QY 147 EVDSETVAYHHEPLGVVQIIIPWFPFLMASKMAFALAAAGNCVVLKPARLTPSVLL 206
DB 182 PADQNYOVHTLHEPIGVAGQIIIPWFPFLMAFANKVGFALACGNTIVLKTAETPLTAFYA 241
QY 207 MEIVGDI-IIPGVNVVNGAGVIGEYLAISKRIAKVAFTGSTEVGQOIMQY-ATONITP 264
DB 242 GKLFLEAGLPVGLNIVSGFATAGALASHMDVDKLAFTGSTDTTKVILGLAANSNLRP 301
QY 265 VTLEIGKSPNIVFADVWDEDAFFOKALE--GFALFAFNOGEVTCPSRALVQESIYER 322
DB 302 VTLEIGKSPNIVFADVWDEDAFFOKALE--GFALFAFNOGEVTCPSRALVQESIYER 322
QY 323 FME-----PAIPPVESIFSGNPLDSVTQMGAGVSHGOETILNVDIGKKEGADVLTGGR 378
DB 355 FVEKSKAPALKRV-----VGDPRKGIHQGQIULKQFEKVMKYIKSGIESNATLEGG 408
QY 379 KLEGELEK-DGYILEPTILFG-ONNMRFVEEIEFGVLAFTTKIMEEALANDTQYGL 436
DB 409 ----DQIGDKGYFTQPTVESNVKIDMLIAQDEIFGPVQSILKFSDDVEVIKRNATKYGL 464
QY 437 CAGVWSRNCNLAYKMGRCIQAGRWNTNCYHAYFAHAFAGYKQSGIGRTHKMMLEHYOO 496
DB 465 AAGVETKRLNLTANFVSNAKAGTIVVNCDFVDAIIPFGYKMGSGNCKEKGYSLNNTYLO 524
QY 497 TKCLLSVYSUKP 508
DB 525 IKA-VVIALNKP 535

RESULT 10
US-10-293-865-20

Sequence 20, Application US/10293865
Publication No. US2004010609A1
GENERAL INFORMATION:
APPLICANT: Nikolau, Basil J
APPLICANT: Wurtele, Eve S
APPLICANT: Oliver, David J
APPLICANT: Behal, Robert
APPLICANT: Schnable, Patrick S
APPLICANT: Ke, Jinsuan
APPLICANT: Johnson, Jerry L
APPLICANT: Allred, Carolyn C
APPLICANT: Fatland, Beth
APPLICANT: Wetz, Tsui-Jung
TITLE OF INVENTION: Materials and Methods for the Alteration of Enzyme and
FILE REFERENCE: 21711
CURRENT APPLICATION NUMBER: US/10293,865
CURRENT FILING DATE: 2002-11-13
PRIOR APPLICATION NUMBER: US 09/344,882
PRIOR FILING DATE: 1999-06-25
PRIOR APPLICATION NUMBER: US 60/040,717
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 38
SOFTWARE: PatentIn Ver. 3.1
SEQ ID NO 20
LENGTH: 538

TYPE: PRT
ORGANISM: Arabidopsis Thaliana
US-10-293-865-20

Query Match 34.4%; Score 918; DB 9; Length 548;
Best Local Similarity 41.1%; Pred. No. 2,96-77;
Matches 202; Conservative 90; Mismatches 170; Indels 40; Gaps 12;

QY 29 TGGHVAADGEYQNLTPVIGQLICRVASSGRKIDIALDAAHKVKDK--WAHITSVQUR 86
DB 62 INGNFVSASCKTPTLPPPTGEVIAHVAEEDAEINRAVKAAPTAFADESPWPKMSAYER 121
QY 87 AAILFKIADRMEOQLLELATAETWNGKPIRETSAADVPLAIDHFRVYFASCIQAQEGGIS 146
DB 122 SRVLLRPADLVKHSSELASLETWNGKPYQOSITAEIPMFARLPFYAGWADKIHGLTI 181
QY 147 EVDSETVAYHHEPLGVVQIIIPWFPFLMASKMAFALAAAGNCVVLKPARLTPSVLL 206
DB 182 PADQNYOVHTLHEPIGVAGQIIIPWFPFLMAFANKVGFALACGNTIVLKTAETPLTAFYA 241
QY 207 MEIVGDI-IIPGVNVVNGAGVIGEYLAISKRIAKVAFTGSTEVGQOIMQY-ATONITP 264
DB 242 GKLFLEAGLPVGLNIVSGFATAGALASHMDVDKLAFTGSTDTTKVILGLAANSNLRP 301
QY 265 VTLEIGKSPNIVFADVWDEDAFFOKALE--GFALFAFNOGEVTCPSRALVQESIYER 322
DB 302 VTLEIGKSPNIVFADVWDEDAFFOKALE--GFALFAFNOGEVTCPSRALVQESIYER 322
QY 323 FME-----PAIPPVESIFSGNPLDSVTQMGAGVSHGOETILNVDIGKKEGADVLTGGR 378
DB 355 FVEKSKAPALKRV-----VGDPRKGIHQGQIULKQFEKVMKYIKSGIESNATLEGG 408
QY 379 KLEGELEK-DGYILEPTILFG-ONNMRFVEEIEFGVLAFTTKIMEEALANDTQYGL 436
DB 409 ----DQIGDKGYFTQPTVESNVKIDMLIAQDEIFGPVQSILKFSDDVEVIKRNATKYGL 464
QY 437 CAGVWSRNCNLAYKMGRCIQAGRWNTNCYHAYFAHAFAGYKQSGIGRTHKMMLEHYOO 496
DB 465 AAGVETKRLNLTANFVSNAKAGTIVVNCDFVDAIIPFGYKMGSGNCKEKGYSLNNTYLO 524
QY 497 TKCLLSVYSUKP 508
DB 525 IKA-VVIALNKP 535

RESULT 11
US-09-961-403-12

Sequence 12, Application US/09961403
Publication No. US20040077589A1
GENERAL INFORMATION:
APPLICANT: HE-STUMPP, HOLGER
APPLICANT: HAEHLER, BERNARD
APPLICANT: KRAETZSCHMAR, JOERN
APPLICANT: KREFT, BERTHOLD
APPLICANT: WINTERBACHER, ELKE
APPLICANT: REICHER, PEDRO
APPLICANT: SCOTT, SIMONE
TITLE OF INVENTION: METHOD FOR IN VITRO DIAGNOSIS OF ENDOMETRIOSIS
FILE REFERENCE: SCH-1789
CURRENT APPLICATION NUMBER: US/09/961,403
CURRENT FILING DATE: 2001-09-25
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 12
LENGTH: 512
TYPE: PRT
ORGANISM: Homo sapiens
US-09-961-403-12

Query Match 33.9%; Score 905; DB 9; Length 512;
Best Local Similarity 39.0%; Pred. No. 4,56-76;
Matches 201; Conservative 88; Mismatches 202; Indels 24; Gaps 12;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd

OM protein - protein search, using sw model

Run on Tue Jun 24, 2003, 10:11:00, Search time 185 sec, 2 seconds
(without alignments)
1780 862 Million cell updates/sec

Title: US-09-830-751-8
Perfect score: 2673
Sequence: 1 MTNPPSAQIKPGVGPPLK ... HYQTKLIVSNHKLGLF 512

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 4569144 seqs, 64473110 residues
Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_Main.*
1: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
9: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
10: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
11: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
12: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
13: /cgn2_6/ptodata/1/paa/US08_COMB.pep.*
14: /cgn2_6/ptodata/1/paa/US09_COMB.pep.*
15: /cgn2_6/ptodata/1/paa/US09_COMB.pep.*
16: /cgn2_6/ptodata/1/paa/US09_COMB.pep.*
17: /cgn2_6/ptodata/1/paa/US09_COMB.pep.*
18: /cgn2_6/ptodata/1/paa/US09_COMB.pep.*
19: /cgn2_6/ptodata/1/paa/US09_COMB.pep.*
20: /cgn2_6/ptodata/1/paa/US09_COMB.pep.*
21: /cgn2_6/ptodata/1/paa/US09_COMB.pep.*
22: /cgn2_6/ptodata/1/paa/US09_COMB.pep.*
23: /cgn2_6/ptodata/1/paa/US09_COMB.pep.*
24: /cgn2_6/ptodata/1/paa/US10_COMB.pep.*
25: /cgn2_6/ptodata/1/paa/US10_COMB.pep.*
26: /cgn2_6/ptodata/1/paa/US10_COMB.pep.*
27: /cgn2_6/ptodata/1/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2668	99.8	512	21	US-09-791-537-8581
2	2668	99.8	512	21	US-09-791-537-8581
3	2668	99.8	512	21	US-09-791-537-8581
4	2668	99.8	512	21	US-09-791-537-8581
5	2668	99.8	512	21	US-09-791-537-8581
6	2668	99.8	512	21	US-09-791-537-8581

7	1917	71.7	495	27	US-60-360-039-15459	Sequence 15459, A
8	1917	71.7	495	27	US-60-360-039-15459	Sequence 15459, A
9	1917	71.7	495	27	US-60-360-039-15459	Sequence 15459, A
10	1917	71.7	495	27	US-60-360-039-15459	Sequence 15459, A
11	1917	71.7	495	27	US-60-360-039-15459	Sequence 15459, A
12	1917	71.7	495	27	US-60-360-039-15459	Sequence 15459, A
13	1917	71.7	495	27	US-60-360-039-15459	Sequence 15459, A
14	1917	71.7	495	27	US-60-360-039-15459	Sequence 15459, A
15	1917	71.7	495	27	US-60-360-039-15459	Sequence 15459, A
16	1917	71.7	495	27	US-60-360-039-15459	Sequence 15459, A
17	1917	71.7	495	27	US-60-360-039-15459	Sequence 15459, A
18	1917	71.7	495	27	US-60-360-039-15459	Sequence 15459, A
19	1917	71.7	495	27	US-60-360-039-15459	Sequence 15459, A
20	1917	71.7	495	27	US-60-360-039-15459	Sequence 15459, A
21	1917	71.7	495	27	US-60-360-039-15459	Sequence 15459, A
22	1917	71.7	495	27	US-60-360-039-15459	Sequence 15459, A
23	1917	71.7	495	27	US-60-360-039-15459	Sequence 15459, A
24	1917	71.7	495	27	US-60-360-039-15459	Sequence 15459, A
25	1917	71.7	495	27	US-60-360-039-15459	Sequence 15459, A
26	1917	71.7	495	27	US-60-360-039-15459	Sequence 15459, A
27	1917	71.7	495	27	US-60-360-039-15459	Sequence 15459, A

ALIGNMENTS

US-09-791-537-8581

Sequence 8581, Application US/09791537

GENERAL INFORMATION:

APPLICANT: Biocom, Inc.

APPLICANT: Biocom, Inc.

APPLICANT: Biocom, Inc.

APPLICANT: Biocom, Inc.

APPLICANT: Biocom, Inc.

APPLICANT: Biocom, Inc.

APPLICANT: Biocom, Inc.

APPLICANT: Biocom, Inc.

APPLICANT: Biocom, Inc.

APPLICANT: Biocom, Inc.

APPLICANT: Biocom, Inc.

APPLICANT: Biocom, Inc.

APPLICANT: Biocom, Inc.

APPLICANT: Biocom, Inc.

APPLICANT: Biocom, Inc.

APPLICANT: Biocom, Inc.

APPLICANT: Biocom, Inc.

APPLICANT: Biocom, Inc.

APPLICANT: Biocom, Inc.

APPLICANT: Biocom, Inc.

APPLICANT: Biocom, Inc.

APPLICANT: Biocom, Inc.

APPLICANT: Biocom, Inc.

APPLICANT: Biocom, Inc.

APPLICANT: Biocom, Inc.

APPLICANT: Biocom, Inc.

APPLICANT: Biocom, Inc.

APPLICANT: Biocom, Inc.

Db 61 KFDIDLALDAHKKVKKWAHTSVQDRAAIIIFKIAHMEANGELIAIAETWONGKPIRETS 120
QY 121 AADVPLAIDHFRYFASCIRAEQGGISEVDSEIVAYHFEPLGVVGGQIIPWNPFLLMASWK 180
Db 121 AADVPLAIDHFRYFASCIRAEQGGISEVDSEIVAYHFEPLGVVGGQIIPWNPFLLMASWK 180
QY 181 MAPALAAGNCVVLKPARLTPLSVLLMEIVGQLLPPGVVNVVNGAGGVIGEYATSKRIA 240
Db 181 MAPALAAGNCVVLKPARLTPLSVLLMEIVGQLLPPGVVNVVNGAGGVIGEYATSKRIA 240
QY 241 KVAFTGSTVGQOIMQYATONIIIPVTILELGGKSPNIVFADVMDEDAEDFKALEGFALFA 300
Db 241 KVAFTGSTVGQOIMQYATONIIIPVTILELGGKSPNIVFADVMDEDAEDFKALEGFALFA 300
QY 301 FNOGEVCTPSRALVQESIVYERPMERAIIRKVESIRSGNPLDSVTOMGADVSHQIETLIN 360
Db 301 FNOGEVCTPSRALVQESIVYERPMERAIIRKVESIRSGNPLDSVTOMGADVSHQIETLIN 360
QY 361 YIDIGKKEGADVLTGGRRKILLEGELKDGYYLEPTILFGONNMRFVQEEIFGVLAVTTFK 420
Db 361 YIDIGKKEGADVLTGGRRKILLEGELKDGYYLEPTILFGONNMRFVQEEIFGVLAVTTFK 420
QY 421 TMEALELANDTOYGLGAGVWSRNGNLAYKMGRIQAGRVWTCYHAYPAHAAPFGYKQS 480
Db 421 TMEALELANDTOYGLGAGVWSRNGNLAYKMGRIQAGRVWTCYHAYPAHAAPFGYKQS 480
QY 481 GIGRETHKMMLEHYQOQTKLLVSYSDKPLGLF 512
Db 481 GIGRETHKMMLEHYQOQTKLLVSYSDKPLGLF 512

RESULT 2

US-09-791-537-9101

: Sequence 9101, Application US/09791537

: GENERAL INFORMATION:

: APPLICANT: Biomimix, Inc.

: APPLICANT: Debe, Derek

: APPLICANT: Danzer, Joseph

: TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS

: TITLE OF INVENTION: METHODS OF USE THEREOF

: FILE REFERENCE: 261/210

: CURRENT APPLICATION NUMBER: US/09/791,537

: CURRENT FILING DATE: 2001-02-22

: NUMBER OF SEQ ID NOS: 153055

: SOFTWARE: PatentIn version 3.0

: SEQ ID NO 9101

: LENGTH: 542

: TYPE: PRT

: ORGANISM: Escherichia coli

US-09-791-537-9101

Query Match 99.8%; Score 2668; DB 21; Length 542,
Best Local Similarity 99.8%; Prod. No. 1.6e 265;
Matches 511; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTNNPSSAQIKPEYGFPLKARYDNFTGGGEWVAPADGEYYQNIIPVTGQLICEVASSG 60
Db 31 MTNNPSSAQIKPEYGFPLKARYDNFTGGGEWVAPADGEYYQNIIPVTGQLICEVASSG 90
QY 61 KRDIIDALDAHKKVKKWAHTSVQDRAAIIIFKIAHMEANGELIAIAETWONGKPIRETS 120
Db 91 KRDIIDALDAHKKVKKWAHTSVQDRAAIIIFKIAHMEANGELIAIAETWONGKPIRETS 150
QY 121 AADVPLAIDHFRYFASCIRAEQGGISEVDSEIVAYHFEPLGVVGGQIIPWNPFLLMASWK 180
Db 151 AADVPLAIDHFRYFASCIRAEQGGISEVDSEIVAYHFEPLGVVGGQIIPWNPFLLMASWK 210
QY 181 MAPALAAGNCVVLKPARLTPLSVLLMEIVGQLLPPGVVNVVNGAGGVIGEYATSKRIA 240
Db 211 MAPALAAGNCVVLKPARLTPLSVLLMEIVGQLLPPGVVNVVNGAGGVIGEYATSKRIA 270
QY 241 KVAFTGSTVGQOIMQYATONIIIPVTILELGGKSPNIVFADVMDEDAEDFKALEGFALFA 300

Db 271 KVAFTGSTVGQOIMQYATONIIIPVTILELGGKSPNIVFADVMDEDAEDFKALEGFALFA 340
QY 301 FNOGEVCTPSRALVQESIVYERPMERAIIRKVESIRSGNPLDSVTOMGADVSHQIETLIN 360
Db 331 FNOGEVCTPSRALVQESIVYERPMERAIIRKVESIRSGNPLDSVTOMGADVSHQIETLIN 390
QY 361 YIDIGKKEGADVLTGGRRKILLEGELKDGYYLEPTILFGONNMRFVQEEIFGVLAVTTFK 420
Db 391 YIDIGKKEGADVLTGGRRKILLEGELKDGYYLEPTILFGONNMRFVQEEIFGVLAVTTFK 450
QY 421 TMEALELANDTOYGLGAGVWSRNGNLAYKMGRIQAGRVWTCYHAYPAHAAPFGYKQS 480
Db 451 TMEALELANDTOYGLGAGVWSRNGNLAYKMGRIQAGRVWTCYHAYPAHAAPFGYKQS 510
QY 481 GIGRETHKMMLEHYQOQTKLLVSYSDKPLGLF 512
Db 511 GIGRETHKMMLEHYQOQTKLLVSYSDKPLGLF 542

RESULT 3

US-09-360-039-23692

: Sequence 23692, Application US/00460049

: GENERAL INFORMATION:

: APPLICANT: Cao, Yongwei

: APPLICANT: Chen, Xianfeng

: APPLICANT: Goldman, Barry S.

: APPLICANT: Hinkle, Gregory J.

: APPLICANT: Slater, Steven C.

: TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

: TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

: FILE REFERENCE: 48-10(52052)A

: CURRENT APPLICATION NUMBER: US/60/360,039

: CURRENT FILING DATE: 2002-02-21

: NUMBER OF SEQ ID NOS: 47374

: SEQ ID NO 23692

: LENGTH: 542

: TYPE: PRT

: ORGANISM: Escherichia coli

US-09-360-039-23692

Query Match 99.8%; Score 2648; DB 27; Length 542;
Best Local Similarity 99.8%; Prod. No. 1.6e 265;
Matches 511; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTNNPSSAQIKPEYGFPLKARYDNFTGGGEWVAPADGEYYQNIIPVTGQLICEVASSG 60
Db 31 MTNNPSSAQIKPEYGFPLKARYDNFTGGGEWVAPADGEYYQNIIPVTGQLICEVASSG 90
QY 61 KRDIIDALDAHKKVKKWAHTSVQDRAAIIIFKIAHMEANGELIAIAETWONGKPIRETS 120
Db 91 KRDIIDALDAHKKVKKWAHTSVQDRAAIIIFKIAHMEANGELIAIAETWONGKPIRETS 150
QY 121 AADVPLAIDHFRYFASCIRAEQGGISEVDSEIVAYHFEPLGVVGGQIIPWNPFLLMASWK 180
Db 151 AADVPLAIDHFRYFASCIRAEQGGISEVDSEIVAYHFEPLGVVGGQIIPWNPFLLMASWK 210
QY 181 MAPALAAGNCVVLKPARLTPLSVLLMEIVGQLLPPGVVNVVNGAGGVIGEYATSKRIA 240
Db 211 MAPALAAGNCVVLKPARLTPLSVLLMEIVGQLLPPGVVNVVNGAGGVIGEYATSKRIA 270
QY 241 KVAFTGSTVGQOIMQYATONIIIPVTILELGGKSPNIVFADVMDEDAEDFKALEGFALFA 300
Db 271 KVAFTGSTVGQOIMQYATONIIIPVTILELGGKSPNIVFADVMDEDAEDFKALEGFALFA 340
QY 301 FNOGEVCTPSRALVQESIVYERPMERAIIRKVESIRSGNPLDSVTOMGADVSHQIETLIN 360
Db 331 FNOGEVCTPSRALVQESIVYERPMERAIIRKVESIRSGNPLDSVTOMGADVSHQIETLIN 390
QY 361 YIDIGKKEGADVLTGGRRKILLEGELKDGYYLEPTILFGONNMRFVQEEIFGVLAVTTFK 420
Db 391 YIDIGKKEGADVLTGGRRKILLEGELKDGYYLEPTILFGONNMRFVQEEIFGVLAVTTFK 450

QY 421 TMEAELELANDIQGLGAGVWSRNLAYKMGRIQAGRYVNTCYHAYPAHAFAFGYKOS 480
 Db 451 TMEAELELANDIQGLGAGVWSRNLAYKMGRIQAGRYVNTCYHAYPAHAFAFGYKOS 510
 QY 481 GIGRETHKMMLEHYQOTKCLLVSYSDKPLGLF 512
 Db 511 GIGRETHKMMLEHYQOTKCLLVSYSDKPLGLF 542

RESULT 4

US-09-252-691-9975
 ; Sequence 9975, Application US/09252691B
 ; GENERAL INFORMATION:
 ; APPLICANT: Keith G. Weinstein et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
 ; FILE REFERENCE: 107196.135
 ; CURRENT APPLICATION NUMBER: US/09/252.691B
 ; CURRENT FILING DATE: 1999-02-18
 ; NUMBER OF SEQ ID NOS: 11324
 ; SEQ ID NO 9975
 ; LENGTH: 546
 ; TYPE: PRT
 ; ORGANISM: Enterobacter cloacae
 US-09-252-691-9975

Query Match 95.88; Score 2560; DB 16; Length 546;
 Best Local Similarity 94.3%; Pred. No. 2.4e-254;
 Matches 483; Conservative 17; Mismatches 12; Indels 0; Gaps 0;

QY 1 MTNPPSAQIKPGYGFPIKLKAPYDNFIFGFWFVADPGFYQNTPTVTGQLCEVASSG 60
 Db 35 MTNPPSSRIQPGYGFPIKLKPYDNFIFGFWFVADPGFYQNTPTVTGQLCEVASSG 94
 QY 61 KPDIDALDAAHKAKKQWQTSVGRRAALIFKIDRFEQNLLELATAETWONGKPIRETS 120
 Db 95 KPDIDALDAAHKAKKQWQTSVGRRAALIFKIDRFEQNLLELATAETWONGKPIRETS 154
 QY 121 AADVPLAIDHFRYFASCIQAGGISEVDSETVAYHFHEPLGVVGGIIPWNPFLMASWK 180
 Db 155 AADVPLAIDHFRYFASCIQAGGISEVDSETVAYHFHEPLGVVGGIIPWNPFLMASWK 214
 QY 181 MAPALAAAGNCVVLKPARLTPLSVLLMEVIGDLPVGVINVNGAGGEGEYLAISKRIA 240
 Db 215 MAPALAAAGNCVVLKPARLTPLSVLLMEVIGDLPVGVINVNGAGGEGEYLAISKRIA 274
 QY 241 KVAFTGSTEVGGQIMQATONIIPTVLELGGKSPNIVFADVMBDEDAFFDKALEGFALFA 300
 Db 275 KVAFTGSTEVGGQIMQATONIIPTVLELGGKSPNIVFADVMBDEDAFFDKALEGFALFA 334
 QY 301 FNQGEVCTCPSPKALVQESIVYERFMEKALRVESIFSPNPLNTQMGAVSHGQLETLN 360
 Db 335 FNQGEVCTCPSPKALVQESIVYERFMEKALRVESIFSPNPLNTQMGAVSHGQLETLN 394
 QY 361 YIDIGKKKGAIVLGGRRKKLIEGELKDYVLEPTILFGQNNMRVFQEEIPGPVLAVTTK 420
 Db 395 YIDIGKKKGAIVLGGRRKKLIEGELKDYVLEPTILFGQNNMRVFQEEIPGPVLAVTTK 454
 QY 421 TMEAELELANDIQGLGAGVWSRNLAYKMGRIQAGRYVNTCYHAYPAHAFAFGYKOS 480
 Db 455 TMEAELELANDIQGLGAGVWSRNLAYKMGRIQAGRYVNTCYHAYPAHAFAFGYKOS 514
 QY 481 GIGRETHKMMLEHYQOTKCLLVSYSDKPLGLF 512
 Db 515 GIGRETHKMMLEHYQOTKCLLVSYSDKPLGLF 546

RESULT 5

US-09-252-691C-9975
 ; Sequence 9975, Application US/09252691C
 ; GENERAL INFORMATION:
 ; APPLICANT: Keith G. Weinstein et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER

; TITLE OF INVENTION: CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.135
 ; CURRENT APPLICATION NUMBER: US/09/252.691C
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,145
 ; PRIOR FILING DATE: 1998-07-24
 ; PRIOR APPLICATION NUMBER: US 60/074,787
 ; PRIOR FILING DATE: 1998-02-18
 ; NUMBER OF SEQ ID NOS: 11326
 ; SEQ ID NO 9975
 ; LENGTH: 546
 ; TYPE: PRT
 ; ORGANISM: Enterobacter cloacae
 US-09-252-691C-9975

Query Match 95.88; Score 2560; DB 16; Length 546;
 Best Local Similarity 94.3%; Pred. No. 2.4e-254;
 Matches 483; Conservative 17; Mismatches 12; Indels 0; Gaps 0;

QY 1 MTNPPSAQIKPGYGFPIKLKAPYDNFIFGFWFVADPGFYQNTPTVTGQLCEVASSG 60
 Db 35 MTNPPSSRIQPGYGFPIKLKPYDNFIFGFWFVADPGFYQNTPTVTGQLCEVASSG 94
 QY 61 KPDIDALDAAHKAKKQWQTSVGRRAALIFKIDRFEQNLLELATAETWONGKPIRETS 120
 Db 95 KPDIDALDAAHKAKKQWQTSVGRRAALIFKIDRFEQNLLELATAETWONGKPIRETS 154
 QY 121 AADVPLAIDHFRYFASCIQAGGISEVDSETVAYHFHEPLGVVGGIIPWNPFLMASWK 180
 Db 155 AADVPLAIDHFRYFASCIQAGGISEVDSETVAYHFHEPLGVVGGIIPWNPFLMASWK 214
 QY 181 MAPALAAAGNCVVLKPARLTPLSVLLMEVIGDLPVGVINVNGAGGEGEYLAISKRIA 274
 Db 215 MAPALAAAGNCVVLKPARLTPLSVLLMEVIGDLPVGVINVNGAGGEGEYLAISKRIA 274
 QY 241 KVAFTGSTEVGGQIMQATONIIPTVLELGGKSPNIVFADVMBDEDAFFDKALEGFALFA 300
 Db 275 KVAFTGSTEVGGQIMQATONIIPTVLELGGKSPNIVFADVMBDEDAFFDKALEGFALFA 334
 QY 301 FNQGEVCTCPSPKALVQESIVYERFMEKALRVESIFSPNPLNTQMGAVSHGQLETLN 360
 Db 335 FNQGEVCTCPSPKALVQESIVYERFMEKALRVESIFSPNPLNTQMGAVSHGQLETLN 394
 QY 361 YIDIGKKKGAIVLGGRRKKLIEGELKDYVLEPTILFGQNNMRVFQEEIPGPVLAVTTK 420
 Db 395 YIDIGKKKGAIVLGGRRKKLIEGELKDYVLEPTILFGQNNMRVFQEEIPGPVLAVTTK 454
 QY 421 TMEAELELANDIQGLGAGVWSRNLAYKMGRIQAGRYVNTCYHAYPAHAFAFGYKOS 480
 Db 455 TMEAELELANDIQGLGAGVWSRNLAYKMGRIQAGRYVNTCYHAYPAHAFAFGYKOS 514
 QY 481 GIGRETHKMMLEHYQOTKCLLVSYSDKPLGLF 512
 Db 515 GIGRETHKMMLEHYQOTKCLLVSYSDKPLGLF 546

RESULT 6

US-60-360-039-17072
 ; Sequence 17072, Application US/60360039
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Chen, Xianfeng
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10,520,52A
 ; CURRENT APPLICATION NUMBER: US/60/360,039
 ; CURRENT FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 17072
 ; LENGTH: 506


```

Db 185 SILVMEVIGDLPPGVNVVNGFGLGAKPLASNPRIAKTFTGTTTGTIMOVASON 244
QY 262 IIPVILEGGKSPNIVFADVWDEEDAFDCKALEGAFAPNQGVEVTCPSRALVQESIYE 321
Db 245 IIPVILEGGKSPNIVFADVWDEEDAFDCKALEGAFAPNQGVEVTCPSRALVQESIYE 304
QY 322 REMEAIKRVKSPNIVFADVWDEEDAFDCKALEGAFAPNQGVEVTCPSRALVQESIYE 381
Db 305 TMEKALKRVAIKOGNPLDPTNVMGAQASSEGLEKILSYIDIGKQEGAEILIGGERNML 364
QY 382 EGELKDGYYLEPTILFGQNNMRVQEEIFGVLAVTTFTKMEAELELANDTQYGLGAGVW 441
Db 365 DGEISDGEYVKTVEFGHKNMVEFQEEIFGVLAVTTFTKMEAELELANDTQYGLGAGVW 424
QY 442 SNGNLAYKMGAGIUGAVWNTNCHYHAYPAHAAPGSGYKUSGIGRETHKMWLFHQOTKCLL 501
Db 425 SDASRLRYMRAIUGAVWNTNCHYHAYPAHAAPGSGYKUSGIGRETHKMWLFHQOTKCLL 484
QY 502 VSYSDKPLGLF 512
Db 485 VSYSPKALGFF 495

RESULT 9
US-60-360-039-16210
: Sequence 16210, Application US/60360039
: GENERAL INFORMATION:
: APPLICANT: Cao, Yongwei
: APPLICANT: Chen, Xianfeng
: APPLICANT: Goldman, Barry S.
: APPLICANT: Hinkle, Gregory J.
: APPLICANT: Slater, Steven C.
: TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
: FILE REFERENCE: 38-10(52052)A
: CURRENT APPLICATION NUMBER: US/60/360,039
: CURRENT FILING DATE: 2002-02-21
: NUMBER OF SEQ ID NOS: 47374
: SEQ ID NO 16210
: LENGTH: 495
: TYPE: PRT
: ORGANISM: Xanthomonas campestris
US-60-360-039-16210

```

```

Query Match 71.7%; Score 1917; DB 27; Length 495;
Best Local Similarity 71.9%; Pred No. 5.4e-188;
Matches 353; Conservative 58, Mismatches 80, Indels 0; Gaps 0;

QY 22 KARYDNFIQGEWVAFADGEYYQNLITFTVTGQLLCEVASGGRKIDLDALDAAHKVKWAHT 81
Db 5 KRYDNFIQGEWVAFADGEYYQNLITFTVTGQLLCEVASGGRKIDLDALDAAHKVKWAHT 64
QY 82 SVQDRAALFKIADRMENLELATAETWNGKPIRETSAADVPLAIDHFRYFASCIQRAQ 141
Db 65 SITDPSNVLLKTIADRIEIANLKLVAESIDNGKPVRETAAIDPLAVDHFYFASCIQRAQ 124
QY 142 EGGISEVDSETVAYHFHEPLGVGQIIPWNEPFLMASWKMALPAAAGNCVVLKPAKLTPL 201
Db 125 EGGISEVDSETVAYHFHEPLGVGQIIPWNEPFLMASWKMALPAAAGNCVVLKPAKLTPL 184
QY 202 SVLLLMEIVGDLPPGVNVVNGFGLGAKPLASNPRIAKTFTGTTTGTIMOVASON 261
Db 185 SILVMEVIGDLPPGVNVVNGFGLGAKPLASNPRIAKTFTGTTTGTIMOVASON 244
QY 262 IIPVILEGGKSPNIVFADVWDEEDAFDCKALEGAFAPNQGVEVTCPSRALVQESIYE 321
Db 245 IIPVILEGGKSPNIVFADVWDEEDAFDCKALEGAFAPNQGVEVTCPSRALVQESIYE 304
QY 322 REMEAIKRVKSPNIVFADVWDEEDAFDCKALEGAFAPNQGVEVTCPSRALVQESIYE 381
Db 305 TMEKALKRVAIKOGNPLDPTNVMGAQASSEGLEKILSYIDIGKQEGAEILIGGERNML 364
QY 382 EGELKDGYYLEPTILFGQNNMRVQEEIFGVLAVTTFTKMEAELELANDTQYGLGAGVW 441
Db 425 SDASRLRYMRAIUGAVWNTNCHYHAYPAHAAPGSGYKUSGIGRETHKMWLFHQOTKCLL 484
QY 502 VSYSDKPLGLF 512
Db 485 VSYSPKALGFF 495

```

```

Db 365 DGEISDGEYVKTVEFGHKNMVEFQEEIFGVLAVTTFTKMEAELELANDTQYGLGAGVW 424
QY 442 SNGNLAYKMGAGIUGAVWNTNCHYHAYPAHAAPGSGYKUSGIGRETHKMWLFHQOTKCLL 501
Db 425 SDASRLRYMRAIUGAVWNTNCHYHAYPAHAAPGSGYKUSGIGRETHKMWLFHQOTKCLL 484
QY 502 VSYSDKPLGLF 512
Db 485 VSYSPKALGFF 495

RESULT 10
US-09-791-537-8001
: Sequence 8001, Application US/09791537
: GENERAL INFORMATION:
: APPLICANT: Bionomix, Inc.
: APPLICANT: Debe, Derek
: APPLICANT: Danzer, Joseph
: TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
: TITLE OF INVENTION: METHODS OF USE THEREOF
: FILE REFERENCE: 261/210
: CURRENT APPLICATION NUMBER: US/09/791,537
: CURRENT FILING DATE: 2001-02-22
: NUMBER OF SEQ ID NOS: 153055
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 8001
: LENGTH: 506
: TYPE: PRT
: ORGANISM: Ralstonia eutropha
US-09-791-537-8001

```

```

Query Match 71.5%; Score 1910; DB 21; Length 506;
Best Local Similarity 71.5%; Pred No. 4e-187;
Matches 351; Conservative 67, Mismatches 73, Indels 0; Gaps 0;

QY 22 KARYDNFIQGEWVAFADGEYYQNLITFTVTGQLLCEVASGGRKIDLDALDAAHKVKWAHT 81
Db 16 KQYENYICGAWVPPAGGEYFESTPTITGKPTFRVRSQQDQVDAALDAHAAKAAWART 75
QY 82 SVQDRAALFKIADRMENLELATAETWNGKPIRETSAADVPLAIDHFRYFASCIQRAQ 141
Db 76 STTERANILNRIADRIEIANLKLVAESIDNGKPVRETAAIDPLAVDHFYFASCIQRAQ 135
QY 142 EGGISEVDSETVAYHFHEPLGVGQIIPWNEPFLMASWKMALPAAAGNCVVLKPAKLTPL 201
Db 136 EGGISEVDSETVAYHFHEPLGVGQIIPWNEPFLMASWKMALPAAAGNCVVLKPAKLTPL 195
QY 202 SVLLLMEIVGDLPPGVNVVNGFGLGAKPLASNPRIAKTFTGTTTGTIMOVASON 261
Db 196 SILVMEVIGDLPPGVNVVNGFGLGAKPLASNPRIAKTFTGTTTGTIMOVASON 255
QY 262 IIPVILEGGKSPNIVFADVWDEEDAFDCKALEGAFAPNQGVEVTCPSRALVQESIYE 321
Db 256 IIPVILEGGKSPNIVFADVWDEEDAFDCKALEGAFAPNQGVEVTCPSRALVQESIYE 315
QY 322 REMEAIKRVKSPNIVFADVWDEEDAFDCKALEGAFAPNQGVEVTCPSRALVQESIYE 381
Db 316 REMEAIKRVKSPNIVFADVWDEEDAFDCKALEGAFAPNQGVEVTCPSRALVQESIYE 375
QY 382 EGELKDGYYLEPTILFGQNNMRVQEEIFGVLAVTTFTKMEAELELANDTQYGLGAGVW 441
Db 376 DGEISDGEYVKTVEFGHKNMVEFQEEIFGVLAVTTFTKMEAELELANDTQYGLGAGVW 435
QY 442 SNGNLAYKMGAGIUGAVWNTNCHYHAYPAHAAPGSGYKUSGIGRETHKMWLFHQOTKCLL 501
Db 446 TMEKALKRVAIKOGNPLDPTNVMGAQASSEGLEKILSYIDIGKQEGAEILIGGERNML 495
QY 502 VSYSDKPLGLF 512
Db 496 VSYSPKALGFF 506

```

RESULT 11

US-60-360-039-12218
 : Sequence 12218, Application US/60360039

: GENERAL INFORMATION:

: APPLICANT: Cao, Yongwei
 : APPLICANT: Chen, Xianfeng
 : APPLICANT: Goldman, Barry S.
 : APPLICANT: Hinkle, Gregory J.
 : APPLICANT: Slater, Steven C.

: TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 : TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

: FILE REFERENCE: 38-10(52052)A

: CURRENT APPLICATION NUMBER: US/60/360,039

: CURRENT FILING DATE: 2002-02-21

: NUMBER OF SEQ. ID NOS: 47374

: SEQ. ID NO 12218

: LENGTH: 505

: TYPE: PRT

: ORGANISM: Mesorhizobium loti

US-60-360-039-12218

Query Match: 71.3%; Score 1905; DB 27; Length 505;

Best Local Similarity 70.3%; Pred. No. 9,66-187;

Matches 352; Conservative 63; Mismatches 84; Indels 2; Gaps 1;

QY 14 EYGFPLK--LKARYDNF IGGEWVAPADGEYVQNLTPVTGQLLCEVASSGKRDIDLALDAA 71

Db 5 EFSRPVAPKDFKRYGNF IGCKWTEPRSGRYFNHSPVQQLCEVARSADDDIEAALDAA 64

QY 72 HKYDKWAHTSVQDRAALEK IADRMQNI ELLEATETWNGKPIRETSAADVPIADHF 131

Db 65 HAADAGRTSVSAERL IINRIADRMENL DLAETWNGKPIRETIVADVPIADHF 124

QY 132 RYFASCIQAEGGISEVDSETVAYHEHPLGVVGGIIPWFFLLMASWKMAPALAAAGNCV 191

Db 125 RYFASVYRGFGSLSDIDDDTVAYHEHPLGVVGGIIPWFFLLMACWKLAALAAAGNCV 184

QY 192 VLKPARUTPLSVLLMEIVGDLPPGVVNVVNGAGVIGEYIATSKRIAKVAFGTGSEVG 251

Db 185 VLKPAEUTPAILLWADLIGDLPPGVVNVVNGAGVIGEYIATSKRIAKVAFGTGTTG 244

QY 252 QQIMQVATONIIPVTLELGKSPNIIVFALVMEEDAFD KALEGFALFAFNGVC(CPS 311

Db 245 RLIMQVATONIIPVTLELGKSPNIIFKQVVAEDDGF D KALEGFVMEALNAGEVC(CPS 304

QY 312 RALVQESIYERFERAI RYVESIRSGNPLDSTOMGAQVSHGQLETIILNYIDIGKKEGAD 371

Db 305 RALIHESIYDRFERAL KRVEAIVQGDPLDPATMICAQASSPQLKILSYIDIGHOGAE 364

QY 372 VLTGGRKLEGEKDCYILEPTILFGQNNMRVFOEEIFGPVLA VTTFKTMEEALELAND 431

Db 365 VLTGGARNVLPGLAGSYVYKPTVFRGHNMKRIFOEEIFGPVVS VTTFKDDDEALSAND 424

QY 432 TOYGLGAVMSRGNLAYKMGRIQAGRWTCNCHYAYFAHAAFGYKSGIGRFTHKMML 491

Db 425 TLYGLCAVWTRCNRAYRFGRAIQAGRWTCNCHYAYFAHAAFGYKSGIGRFTHKMML 484

QY 492 EHYQQTCKLLVSYSDKPLGLF 512

Db 485 DHYQQTCKMLVSYSPKKLGLF 505

RESULT 12

US-60-360-039-4614

: Sequence 4614, Application US/6060039

: GENERAL INFORMATION:

: APPLICANT: Cao, Yongwei
 : APPLICANT: Chen, Xianfeng
 : APPLICANT: Goldman, Barry S.
 : APPLICANT: Hinkle, Gregory J.
 : APPLICANT: Slater, Steven C.

: TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 : TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

: FILE REFERENCE: 38-10(52052)A

: CURRENT APPLICATION NUMBER: US/60/360,039

: CURRENT FILING DATE: 2002-02-21

: NUMBER OF SEQ. ID NOS: 47374

: SEQ. ID NO 4614

: LENGTH: 506

: TYPE: PRT

: ORGANISM: Burkholderia fungorum

US-60-360-039-4614

Query Match:

Best Local Similarity 70.6%; Pred. No. 130-186;

Matches 350; Conservative 62; Mismatches 82; Indels 2; Gaps 1;

QY 17 FPLKIRARYDNFTGTFWVAFAAGFYVQNLTPVTGQLLCEVASSGKRDIDLALDAAHKVVD 76

Db 13 FP--YKKQVANTFGGEWVAPADGEYVQNLTPVTGQLLCEVASSGKRDIDLALDAAHRAKT 70

QY 77 KWAHTSVQDRAALEK IADRMQNI ELLEATETWNGKPIRETSAADVPIADHFRYFAS 146

Db 71 AWKTSAAUPANINPIADRMQNI ELLEATETWNGKPIRETSAADVPIADHFRYFAG 140

QY 137 CIRAQEGGISEVDSETVAYHEHPLGVVGGIIPWFFLLMASWKMAPALAAAGNVVILKPA 196

Db 131 AVRAQEGSLSEIDHDITVAYHEHPLGVVGGIIPWFFLLMAVWKMAPALAAAGNVVILKPA 190

QY 197 PTPPLSVLLMEIVGDLPPGVVNVVNGAGVIGEYIATSKRIAKVAFGTGSEVG(SQIMQ 256

Db 191 EQTPASILVVELIQLLPPGVVNVVNGAGVIGEYIATSKRIAKVAFGTGSEVG(SQIMQ 250

QY 257 YATONITPVTLELGKSPNIIVFALVMEEDAFD KALEGFALFAFNGVTTOTSPALVQ 416

Db 251 YASONTIPVTLELGKSPNIIFKQVVAEDDGF D KALEGFAMALNAGEVTTOTSPALVQ 410

QY 317 ESTYEFMFEPALPPVESIHSCHLESVTCMAV--SPELEETILNYIDIGKKEGAD 476

Db 311 EKLYDEMERALKRVAATQGHPLDPTKIMICAQASSPQLKILSYIDIGHOGAE 470

QY 377 RKLLGEELKDGYYLEPTILFGQNNMRVFOEEIFGPVLA VTTFKTMEEALELAND 436

Db 371 ERNALGGELSKGYVYKPTVFRGHNMKRIFOEEIFGPVVS VTTFKTMEEALELAND 430

QY 437 GACVWSPNMIAYKMGRIQAGRWTCNCHYAYFAHAAFGYKSGIGRFTHKMML 496

Db 431 GAGVWTRDGTTRAYRFGRAIQAGRWTCNCHYAYFAHAAFGYKSGIGRFTHKMML 490

QY 497 TKCLLVSYSDKPLGLF 512

Db 491 TKNLVSYSDKPLGLF 506

RESULT 13

US-60-360-039-7371

: Sequence 7371, Application US/6060049

: GENERAL INFORMATION:

: APPLICANT: Cao, Yongwei
 : APPLICANT: Chen, Xianfeng
 : APPLICANT: Goldman, Barry S.
 : APPLICANT: Hinkle, Gregory J.
 : APPLICANT: Slater, Steven C.

: TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 : TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

: FILE REFERENCE: 38-10(52052)A

: CURRENT APPLICATION NUMBER: US/60/360,039

: CURRENT FILING DATE: 2002-02-21

: NUMBER OF SEQ. ID NOS: 47374

: SEQ. ID NO 7371

: LENGTH: 506

: TYPE: PRT

: ORGANISM: Burkholderia cepacia

US-60-360-039-7371

Query Match

70.9%; Score 1894; DB 27; Length 506;

Best Local Similarity 70.68; Pred. No. 13e-185;
Matches 350; Conservative 62; Mismatches 82; Indels 2; Gaps 1;

QY 17 PPLKLYDNFNGEYVADGGEYONLTPVTGQLCEVASSGKRDIDLALDAAHKVD 76
Db 13 PP--YKKQYANFVGGGKVGGEYFONVSPTGEAFTSPRSPADVPALDAAHKPT 70
QY 77 KWHTSVQDRAALFKIADRMENONLELATAETWNGKPIRETSAADVPALDHPRYFAS 136
Db 71 AWGTSAADRANILNRADMEANLQRLVAETIDNGKPLRETAAIDPLADHPRYFAG 130
QY 137 CIRAEGGISEVDSETVAYHFHEPLGVVGGQIIPWNEPPLMASMKMAPALAGNCVVLKPA 196
Db 131 AVRAQEGSISFIDHPTVAYHFHEPLGVVGGQIIPWNEPPLMASMKMAPALAGNCVVLKPA 190
QY 197 RITPLSVLLMEIVGDLPLPGSVNVVNGAGSVIGEYLAATSKRIAKVAFVGTSTEVGQIMQ 256
Db 191 EQIFASILVVELLDLLPPVNLNVNGSPLEAKKPIASSKPIAKIAFTSETTTPGLIMQ 250
QY 257 YATGNIIPVLELGGSKSPNIVFAIWMLEKFAFFKAIKQFALPAHQREVCPCSPALVQ 316
Db 251 YASQNIIPVLELGGKSPNIFADVMNEDDSFEDKALEGFAMALNQGEVCTCPSKVLID 310
QY 317 ESIVEREMERAIRVESIRSGNPLDSVYTMGAOVSHGQLETLINYYIDICKGADVLFG 376
Db 311 EKIVDEREMERALKVAAITQCHPDKIMIGASASQELKILSYVDLKGQEGACELGG 370
QY 377 PKLLEGLKGGYVLEPTILPGGNMMPVGFRTFGVPLAVTTEKTMEEALELANDTQGL 436
Db 371 ERNALGGLSKGYVTKPTVFRGNMKNRIFOEEIFGPVSVTTFRNEEEALELANDTQGL 430
QY 437 GAGVMSRGNLAYKMGRIQAGRVWNTNCHAYPAHAAGFGYKQSGIGRETHKMLLHYQO 496
Db 431 GAGVMTDGTATYAFEGFIQAGRVWNTNCHAYPAHAAGFGYKQSGIGRETHKMLLHYQO 490
QY 497 TKLLVSYSDKPLGLF 512
Db 491 TKLLVSYSDKPLGLF 506

RESULT 14

US-60-360-039-11572
; Sequence 11572, Application US/60360039
; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)A
; CURRENT APPLICATION NUMBER: US/60/360,039

; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 11572

; LENGTH: 495

; TYPE: PRT

; ORGANISM: Agrobacterium tumefaciens

US-60-360-039-11572

Query Match 70.78; Score 1890; LR 27; Length 495;

Best Local Similarity 69.78; Pred. No. 33e-185;

Matches 345; Conservative 68; Mismatches 80; Indels 2; Gaps 1;

QY 18 PPLKLYDNFNGEYVADGGEYONLTPVTGQLCEVASSGKRDIDLALDAAHKVD 77
Db 3 PPLK--YCNVIGGKWKPKSPYMYNI SPVSHKICEVPPSPVASHIIFALDAAHKAPK 80
QY 78 KWHTSVQDRAALFKIADRMENONLELATAETWNGKPIRETSAADVPALDHPRYFAS 137
Db 61 WKTSITFRSNIIPVLELGGKSPNIFADVMNEDDSFEDKALEGFAMALNQGEVCTCPSKVLID 120

QY 138 IRAOEGGISEVDSETVAYHFHEPLGVVGGQIIPWNEPPLMASMKMAPALAGNCVVLKPAR 197
Db 121 IRAOEGTIGEDNDTVAYHFHEPLGVVGGQIIPWNEPPLMASMKMAPALAGNCVVLKPAE 180
QY 198 LTPLSVLLMEIVGDLPLPGSVNVVNGAGSVIGEYLAATSKRIAKVAFVGTSTEVGQIMQ 257
Db 181 QTPASILVVELLDLLPPGVLNVNGTGLEAGKPLAQSNRIAKIAFTGSTSVGKEIMRY 240
QY 258 ATQNIIPVLELGGKSPNIVFAIWMLEKFAFFKAIKQFALPAHQREVCPCSPALVQ 317
Db 241 AADNVNISIIEGLGKSPNIFADVMNEDDSFEDKALEGFAMALNQGEVCTCPSKVLID 300
QY 318 SIYEPFEMERAIRVESIRSGNPLDSVYTMGAOVSHGQLETLINYYIDICKGADVLFG 377
Db 301 SIYDPFEMERAIRVESIRSGNPLDSVYTMGAOVSHGQLETLINYYIDICKGADVLFG 360
QY 378 PKLLEGLKGGYVLEPTILPGGNMMPVGFRTFGVPLAVTTEKTMEEALELANDTQGL 437
Db 361 PKLTGDLKGGYVLEPTILPGGNMMPVGFRTFGVPLAVTTEKTMEEALELANDTQGL 420
QY 438 AGVMSRGNLAYKMGRIQAGRVWNTNCHAYPAHAAGFGYKQSGIGRETHKMLLHYQO 497
Db 421 AGVMSRGNLAYKMGRIQAGRVWNTNCHAYPAHAAGFGYKQSGIGRETHKMLLHYQO 480
QY 498 KCLLVSYSDKPLGLF 512
Db 481 KCLLVSYSDKPLGLF 495

RESULT 15

US-09-791-537-31772

; Sequence 31772, Application US/09791537

; GENERAL INFORMATION:

; APPLICANT: Bionomix, Inc.

; APPLICANT: Debe, Derek

; APPLICANT: Danzer, Joseph

; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME

; FILE REFERENCE: 261/210

; CURRENT APPLICATION NUMBER: US/09/791,537

; CURRENT FILING DATE: 2001-02-22

; NUMBER OF SEQ ID NOS: 153055

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 31772

; LENGTH: 505

; TYPE: PRT

; ORGANISM: Xanthobacter autotrophicus

US-09-791-537-31772

Query Match 70.38; Score 1879; LR 21; Length 505;

Best Local Similarity 69.28; Pred. No. 4.7e-184;

Matches 353; Conservative 58; Mismatches 93; Indels 6; Gaps 1;

QY 3 NPESAOIKPEYGFPIKIKAPYINPTGGHVAIAPCEYKNIIVYVQALLTEVASSKRP 62
Db 2 NKPEIATKQSP-----KAPGNFTGGKFPVPGSPYFNTSPVTGKPTTAPSDAD 55
QY 63 DIDLDAAHKVKDKWAHTSVQDRMAILEKIDRMEONLELATAETWNGKPIRETSA 122
Db 56 DIEKALDAHAARKDAWGKTSAAVPAITLINPADRMEENLDLLALAEATWNGKPIRET 115
QY 114 IYVFLAIDHPRFASCTAAGEDTSFVDSSETVAFHFEPGLVAVQIIPWNEPPLMASMKMA 182
Db 116 LMPALIDHPRFAYFAVFAVAGAGAGISEIHHPHIVAYHFHEPLGVVGGQIIPWNEPPLMASMKMA 175
QY 184 PALAAGNCVVLKPARLPLSVLLMEIVGDLPLPGSVNVVNGAGSVIGEYLAATSKRIAKV 242
Db 176 PALAAGNCVVLKPARLPLSVLLMEIVGDLPLPGSVNVVNGAGSVIGEYLAATSKRIAKV 235
QY 243 APTGSGIEVGQIMQYATQNIIPVLELGGKSPNIFADVMNEDDSFEDKALEGFAMALNQ 302
Db 246 APTGSETTGRILMQYASQNLIPVLELGGKSPNIFADVMNEDDSFEDKALEGFAMALNQ 295

QY 303 QGEVTCPSRALVOESIYERFMERAIIRRVESIRSGNPLDSVTQMGAQVSHGOLETILAVI 362
 Db 296 QGEVTCPSRALVOESIYERFMERAIIRRVESIRSGNPLDSVTQMGAQVSHGOLETILAVI 355
 QY 363 DIGKKPGADVITGGPKKLLPCELKGGYILEPTILFQNNMNVFQEEIFGPVLAVTIFRTM 422
 Db 356 DIGKQGAELVIGGERNTFGCDLAGGYVVKPTVFKGHNMRIQBEIFGPVVSVTFTUD 415
 QY 423 EEALELANDTOYGLGAGVWSRGNLAYKMGRTQAGRVWTCYHAYPAHAAPGGYKQSGI 482
 Db 416 AEALHIANDTLGLGA:VWTFQ:NRAYPRFPATQAGVWTCYHAYPAHAAPGGYKQSGI 475
 QY 483 GRETHKMMLEHYQOTKCLJVSYSKPIGLF 512
 Db 476 GRENHKMMLDHYQOTKMMLVSYSPKKLGF 505

Search completed: June 24, 2003, 10:30:44
 Job time : 188.362 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:14:35 ; Search time 55.428 Seconds
(without alignments)
2375.712 Million cell updates/sec

Title: US-09-830-751-8
Perfect score: 2673
Sequence: 1 MTNPPSAQIKPGYGFPIK HYQTKTIVSVSHKPLDIF 512

Scoring table BLASTSUM62

Gapop 10 0, Gapext 0 5

Searched: 1171708 seqs, 257149365 residues

Total number of hits satisfying chosen parameters: 1171708

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_AA_New:*

1: /cgn2_6/ptodata/2/paa/FCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2673	100	0	512	US-09-830-751-8
2	2668	99	8	542	US-10-282-122A-43103
3	2668	99	8	542	US-10-282-122A-43103
4	2589	96	9	512	US-10-282-122A-75409
5	2567	96	0	534	US-10-282-122A-56331
6	2560	95	8	546	US-10-417-886-9975
7	2054	76	8	456	US-10-282-122A-72955
8	1918	71	8	506	US-10-369-493-17072
9	1917	71	7	495	US-10-369-493-15459
10	1917	71	7	495	US-10-369-493-15827
11	1917	71	7	495	US-10-369-493-16210
12	1915	71	6	567	US-10-282-122A-50457
13	1905	71	3	505	US-10-369-493-12218
14	1894	70	9	505	US-10-282-122A-49350
15	1894	70	9	506	US-10-369-493-4614
16	1894	70	9	506	US-10-369-493-7371
17	1890	70	7	445	US-10-369-493-11572
18	1878	70	3	506	US-10-369-493-17140
19	1870	70	0	504	US-10-369-493-19271
20	1868	69	9	493	US-10-369-493-4608
21	1868	69	3	493	US-10-369-493-7365
22	1868	69	4	505	US-10-282-122A-44659
23	1856	69	4	495	US-10-369-493-17869
24	1853	69	3	508	US-10-282-122A-47513
25	1849	69	2	515	US-10-369-493-23469
26	1847	69	1	492	US-10-469-493-20815

27	1839	68.8	507	6	US-10-282-122A-61783	Sequence 61783, A
28	1827.5	68.4	503	6	US-10-282-122A-44658	Sequence 44658, A
29	1823.5	68.2	514	6	US-10-417-652-6997	Sequence 6997, Ap
30	1821	68.1	494	6	US-10-369-493-21741	Sequence 21741, A
31	1819	68.1	505	6	US-10-282-122A-68054	Sequence 68054, A
32	1817	68.0	503	6	US-10-282-122A-63180	Sequence 63180, A
33	1808	67.6	527	6	US-10-282-122A-51390	Sequence 51390, A
34	1802	67.4	505	6	US-10-282-122A-65886	Sequence 65886, A
35	1798	67.3	507	6	US-10-282-122A-62787	Sequence 62787, A
36	1794	67.4	507	6	US-10-282-122A-64382	Sequence 64382, A
37	1795	67.2	682	6	US-10-366-683-31897	Sequence 31897, A
38	1795	67.2	682	6	US-10-419-128-31897	Sequence 31897, A
39	1791	67.0	505	6	US-10-282-122A-65344	Sequence 65344, A
40	1791	67.0	506	6	US-10-282-122A-77335	Sequence 77335, A
41	1783	66.7	506	6	US-10-369-493-8796	Sequence 8796, Ap
42	1773	66.3	506	6	US-10-369-493-14000	Sequence 14000, A
43	1770	66.2	507	6	US-10-156-761-14706	Sequence 14706, A
44	1759	65.8	534	6	US-10-446-203-10148	Sequence 10148, A
45	1758	65.8	506	6	US-10-282-122A-60172	Sequence 60172, A

ALIGNMENTS

RESULT 1

US-09-830-751-8
; Sequence 8, Application US/09830751
; GENERAL INFORMATION:
; APPLICANT: Suthers, Patrick F.
; TITLE OF INVENTION: Production of 3-Hydroxypropionic Acid in Recombinant
; FILE REFERENCE: 960296, 96617
; CURRENT APPLICATION NUMBER: US/09/830,751
; CURRENT FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 65/151,440
; PRIOR FILING DATE: 1999-08-30
; PRIOR APPLICATION NUMBER: PCT/US00/23878
; PRIOR FILING DATE: 2000-08-30
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 512
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-830-751-8

Query Match	100.0%	Score: 2674	DB: 5	Length: 512
Best Local Similarity	100.0%	Prod. No.: 1.76-236	Mismatches: 0	Gaps: 0
Matches: 512	Conservative: 0	Mismatches: 0	Indeles: 0	Gaps: 0
QY	1	MTNPPSAQIKPGYGFPIKAKYUNF	IGGEWVAFAKSEYUNLTPVIGQLCEVASSG	60
DB	1	MTNPPSAQIKPGYGFPIKAKYUNF	IGGEWVAFAKSEYUNLTPVIGQLCEVASSG	60
QY	61	KRDTDLALDAHKVKYKWAHTSVQDRAA	ILFKIADMEQNLLELLAIAFWNCKPIFETS	120
DB	61	KRDTDLALDAHKVKYKWAHTSVQDRAA	ILFKIADMEQNLLELLAIAFWNCKPIFETS	120
QY	121	AADYPLAIDHRYFASCIQAEGGISEVD	SETVAYHFHPLGVVGQIIPNPFLLMASWK	180
DB	121	AADYPLAIDHRYFASCIQAEGGISEVD	SETVAYHFHPLGVVGQIIPNPFLLMASWK	180
QY	181	MAPAAAGNCVVLKPAFTPLSVLLLM	EIVEDLIPGVVNVVNGAGGVIDEYATSEPTA	240
DB	181	MAPAAAGNCVVLKPAFTPLSVLLLM	EIVEDLIPGVVNVVNGAGGVIDEYATSEPTA	240
QY	241	KVAFIGSTEVCQIMQYATQNIIPVILE	LGGKSNIVFAIVMDEEHAFFKALLEGALFA	300
DB	241	KVAFIGSTEVCQIMQYATQNIIPVILE	LGGKSNIVFAIVMDEEHAFFKALLEGALFA	300
QY	301	FNAGSEVCTSPALVHSEIVFERMFP	PAIPVESIRSDNPLDSVTLDMACVSHQIETILN	450
DB	301	FNAGSEVCTSPALVHSEIVFERMFP	PAIPVESIRSDNPLDSVTLDMACVSHQIETILN	450

Db 301 FNQGEVCTCPSRALVQHSIVYERPMERAIIRRVESIRSNPLDSVTOMGAOVSHGQLETTLN 360
 QY 361 YIDIGKKEGADVLTGSRKKLLFGLKDGYYLEPTILFGONNMRVFEELFGVLAATTFFK 420
 Db 361 YIDIGKKEGADVLTGSRKKLLFGLKDGYYLEPTILFGONNMRVFEELFGVLAATTFFK 420
 QY 421 TMEALELANTQYGLGAGWMSRNGNLAYKMGRTQAGRVWTCYHAYPAHAAGFGYKQS 480
 Db 421 TMEALELANTQYGLGAGWMSRNGNLAYKMGRTQAGRVWTCYHAYPAHAAGFGYKQS 480
 QY 481 GIGRETHKMMLEHYQOTKCLLVSYSDKPLGLF 512
 Db 481 GIGRETHKMMLEHYQOTKCLLVSYSDKPLGLF 512

RESULT 2

US-10-282-122A-43103
 ; Sequence 43103, Application US/10282122A

GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Gdout
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 43103

; LENGTH: 542

; TYPE: PRT

; ORGANISM: Escherichia coli

US-10-282-122A-43103

Query Match

Best Local Similarity 99.8%; Score 2668; DB 6; Length 542;

Matches 511; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTNNPSSAQIKPCEYGFPLKARYDNFICGGEHWAPADGEYYQNLTPVTGQLLCEVASSG 60
 Db 31 MTNNPSSAQIKPCEYGFPLKARYDNFICGGEHWAPADGEYYQNLTPVTGQLLCEVASSG 90
 QY 61 KRDIIDLALDAAHKVKDKWAHTSVQDRAAILFKIADRMQNLLELLATAFTWNGKPIRETS 120

Db 91 KRDIIDLALDAAHKVKDKWAHTSVQDRAAILFKIADRMQNLLELLATAFTWNGKPIRETS 150
 QY 121 AADVPLAIDHFRYFASCIRAQEGGISEVUSSETVAVYHFHEPLGVCVQIITWNFPLLMASWK 180
 Db 151 AADVPLAIDHFRYFASCIRAQEGGISEVUSSETVAVYHFHEPLGVCVQIITWNFPLLMASWK 210
 QY 181 MAVALAAGNCVILKPARLTPLSVLLIMEIVGDLDPGVVNVVNGAGVIGEYLATSKRIA 240
 Db 211 MAPALAAGNCVILKPARLTPLSVLLIMEIVGDLDPGVVNVVNGAGVIGEYLATSKRIA 270
 QY 241 KVAFTGSTEVGGQIMQYATONIIPTVLELGCGSPNIVFADVMDEEDAFEDKALGCFALFA 400
 Db 271 KVAFTGSTEVGGQIMQYATONIIPTVLELGCGSPNIVFADVMDEEDAFEDKALGCFALFA 440
 QY 301 FNQGEVCTCPSRALVQHSIVYERPMERAIIRRVESIRSNPLDSVTOMGAOVSHGQLETTLN 460
 Db 331 FNQGEVCTCPSRALVQHSIVYERPMERAIIRRVESIRSNPLDSVTOMGAOVSHGQLETTLN 490
 QY 361 YIDIGKKEGADVLTGSRKKLLFGLKDGYYLEPTILFGONNMRVFEELFGVLAATTFFK 420
 Db 391 YIDIGKKEGADVLTGSRKKLLFGLKDGYYLEPTILFGONNMRVFEELFGVLAATTFFK 450
 QY 421 TMEFALELANDTOYGLGAGWMSRNGNLAYKMGRTQAGRVWTCYHAYPAHAAGFGYKQS 480
 Db 451 TMEFALELANDTOYGLGAGWMSRNGNLAYKMGRTQAGRVWTCYHAYPAHAAGFGYKQS 510
 QY 481 GIGRETHKMMLEHYQOTKCLLVSYSDKPLGLF 512
 Db 511 GIGRETHKMMLEHYQOTKCLLVSYSDKPLGLF 542

RESULT 3

US 10 369 493 23602

; Sequence 23602, Application US/10-69493

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: PLANTS WITH IMPROVED PROPERTIES

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/160,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 23602

; LENGTH: 542

; TYPE: PRT

; ORGANISM: Escherichia coli

US-10-469-493-23602

Query Match

Best Local Similarity 99.8%; Score 2668; DB 6; Length 542;

Matches 511; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTNNPSSAQIKPCEYGFPLKARYDNFICGGEHWAPADGEYYQNLTPVTGQLLCEVASSG 60
 Db 31 MTNNPSSAQIKPCEYGFPLKARYDNFICGGEHWAPADGEYYQNLTPVTGQLLCEVASSG 90
 QY 61 KRDIIDLALDAAHKVKDKWAHTSVQDRAAILFKIADRMQNLLELLATAFTWNGKPIRETS 120
 Db 91 KRDIIDLALDAAHKVKDKWAHTSVQDRAAILFKIADRMQNLLELLATAFTWNGKPIRETS 150
 QY 121 AADVPLAIDHFRYFASCIRAQEGGISEVUSSETVAVYHFHEPLGVCVQIITWNFPLLMASWK 180
 Db 151 AADVPLAIDHFRYFASCIRAQEGGISEVUSSETVAVYHFHEPLGVCVQIITWNFPLLMASWK 210
 QY 181 MAPALAAGNCVILKPARLTPLSVLLIMEIVGDLDPGVVNVVNGAGVIGEYLATSKRIA 240

Db 211 MAPALAGNCVVLKAP:TPISVILMEIVCDLLPFGVNVVWVNCAGGVCEYLATSKRIA 270
QY 241 KVAFTGSTEVGQOIMQYATONIPVTLGKGSFNIVFAVUMJEDAFEDKALEGALFA 400
Db 271 KVAFTGSTEVGQOIMQYATONIPVTLGKGSFNIVFAVUMJEDAFEDKALEGALFA 330
QY 301 FNOGEVCTCPSPALVQESIVYERFERMERAIRPVESIPSGNPLDSVTQMGAGVSHGQLETTIN 360
Db 331 FNOGEVCTCPSPALVQESIVYERFERMERAIRPVESIPSGNPLDSVTQMGAGVSHGQLETTIN 390
QY 361 YIDIGKKGADVLIGGRKKLLEBELKDGYYLEPIILFGQNNMVFQBEIFGPPVLAVTTTK 420
Db 391 YIDIGKKGADVLIGGRKKLLEBELKDGYYLEPIILFGQNNMVFQBEIFGPPVLAVTTTK 450
QY 421 TMEAELELANDTQGLGAGVWSRNGNLAYKMGKGLAGKRWVINCYHAYFAHAAFGGYKUS 480
Db 451 TMEAELELANDTQGLGAGVWSRNGNLAYKMGKGLAGKRWVINCYHAYFAHAAFGGYKUS 510
QY 481 GIGRETHKMMLHYQOTKCLLVSYSDKPLGLF 512
Db 511 GIGRETHKMMLHYQOTKCLLVSYSDKPLGLF 542

RESULT 4

US-10-282-122A-75409
; Sequence 75409, Application US/10282122A
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsuo
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,408
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PAIR.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 75409
; LENGTH: 512
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-10-282-122A-75409

Query Match

96 98: Score: 2580; DB 6; Length: 512;

Best Local Similarity 95.7%, Fred. No. 8.8e-229;
Matches 490; Conservative 14; Mismatches 8; Indels 0; Gaps 0;
QY 1 MINPPSAQIKPGYGFPLKAPYUNFISGWSWAPAKSFYQUNLIPVTGJLLCEVASSG 60
Db 1 MTNPPSTRIGPSEVGYPLKAPYUNFISGWSWAPAKSFYQUNLIPVTGJLLCEVASSG 60
QY 61 KPDIDLDAAHKVKYQWHAITSVQDPAAIILPKTADRMFQNLFIILATAETWNGKPTPTTS 120
Db 61 KKDIDLDAAHKARKWAHTSVQDRAAIIILFKIADRMQNLLELLATAETWNGKPTPTTS 120
QY 121 AADVPLAIDHFRYFASCIKAGEGGISEVDSETVAYHHEPLGVVGQIIPWNEPFLLMASNK 180
Db 121 AADIPLAIDHFRYFASCIKAGEGGISEVDSETVAYHHEPLGVVGQIIPWNEPFLLMASNK 180
QY 181 MAPALAGNCVVLKAP:TPISVILMEIVCDLLPFGVNVVWVNCAGGVCEYLATSKRIA 240
Db 181 MAPALAGNCVVLKAP:TPISVILMEIVCDLLPFGVNVVWVNCAGGVCEYLATSKRIA 240
QY 241 KVAFTGSTEVGQOIMQYATONIPVTLGKGSFNIVFAVUMJEDAFEDKALEGALFA 300
Db 241 KVAFTGSTEVGQOIMQYATONIPVTLGKGSFNIVFAVUMJEDAFEDKALEGALFA 300
QY 301 FNOGEVCTCPSPALVQESIVYERFERMERAIRPVESIPSGNPLDSVTQMGAGVSHGQLETTIN 360
Db 301 FNOGEVCTCPSPALVQESIVYERFERMERAIRPVESIPSGNPLDSVTQMGAGVSHGQLETTIN 360
QY 361 YIDIGKKGADVLIGGRKKLLEBELKDGYYLEPIILFGQNNMVFQBEIFGPPVLAVTTTK 420
Db 361 YIDIGKKGADVLIGGRKKLLEBELKDGYYLEPIILFGQNNMVFQBEIFGPPVLAVTTTK 420
QY 421 TMEAELELANDTQGLGAGVWSRNGNLAYKMGKGLAGKRWVINCYHAYFAHAAFGGYKUS 480
Db 421 TMEAELELANDTQGLGAGVWSRNGNLAYKMGKGLAGKRWVINCYHAYFAHAAFGGYKUS 480
QY 481 GIGRETHKMMLHYQOTKCLLVSYSDKPLGLF 512
Db 481 GIGRETHKMMLHYQOTKCLLVSYSDKPLGLF 512

RESULT 5

US-10-282-122A-56331
; Sequence 56331, Application US/10282122A
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsuo
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625

```

; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,408
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56331
; LENGTH: 534
; TYPE: PRT
; ORGANISM: Enterobacter cloacae
US-10-282-122A-56331

Query Match          96.08; Score 2567; DB 6; Length 534;
Best Local Similarity 94.5%; Pred. No. 1e-226;
Matches 484; Conservative 17; Mismatches 11; Indels 0; Gaps 0;

QY 1 MTNPPSAOIKPGEGFPLKIKARYDNFTGGWVAFADGCEYQNTPTVTGOLLCEVASSG 60
DB 23 MTNPPSSRIQPGEGFPLKIKARYDNFTGGWVADVDGGEYSNLTPTVTGOLCEIASSG 82
QY 61 KRDIIDLALDAAHKVKOKWAHTSVQDRAAIIKFKIADRMENLELLATAETWNGKPIRETS 120
DB 83 KRDIIDLALDAAHKAKDKWGHTSVQDRAAIIKFKIADRMENLELLATAETWNGKPIRETM 142
QY 121 AADVPLAIDHFRYFASCIRAQEGGISEVDSEIVAYHFHETGVVQIIPWNPFLMASWK 180
DB 143 AADVPLAIDHFRYFASCIRAQEGGISEVDSDIVAYHFHETGVVQIIPWNPFLMASWK 202
QY 181 MAPALAGNCVILKPARLTPLSVLLMEIVGDLPPGVNVVNGAGGVIGEYLATSKRIA 240
DB 203 MAPALAGNCVILKPARLTPLSVLLMEIVGDLPPGVNVVNGAGGELGEYLATSKRIA 262
QY 241 KVAFTGSTGVQGIIMQYATQNIIPVTLELGGKSPNIVADVMDDEDAFDDKALEGALFA 300
DB 263 KVAFTGSTGVQGIIMQYATQNIIPVTLELGGKSPNIVADVMDDEDAFDDKALEGALFA 322
QY 301 FNOGEVCTCPSRALVQESIVEREMERAIIRVESIRSGNPLDSVTOMGAUVSHGULETILN 360
DB 323 FNOGEVCTCPSRALVQESIVEREMERAIIRVESIRSGNPLDSVTOMGAUVSHGULETILN 382
QY 361 YIDIGKKEGADVLTGGRRKILLEGELKDGYYLEPTILFGONNRRVFOEIEFGVLAATTEK 420
DB 383 YIDIGKKEGADVLTGGRRKILLEGELKDGYYLEPTILFGONNRRVFOEIEFGVLAATTEK 442
QY 421 TMEALELANDTOYGLCAGVWSNPNLAYKMRGICQAGVWNTNCHAYPAHAAPGCKQS 480
DB 443 TMEALELANDTOYGLCAGVWSNPNLAYKMRGICQAGVWNTNCHAYPAHAAPGCKQS 502
QY 481 GIGRETHKMMLEHYQQTCKLLVSYSOKPLGLF 512
DB 503 GIGRETHKMMLEHYQQTCKLLVSYSOKPLGLF 534

RESULT 6
US-10-417-886-9975
; Sequence 9975, Application US/10417886
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstock et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/10417,886
; CURRENT FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: US/09/252,691C
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/034,145
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: US 60/074,787
; PRIOR FILING DATE: 1998-02-18

```

```

; NUMBER OF SEQ ID NOS: 11326
; SEQ ID NO 9975
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Enterobacter cloacae
US-10-417-886-9975

Query Match          95.88; Score 2560; DB 6; Length 546;
Best Local Similarity 94.4%; Pred. No. 4.6e-226;
Matches 484; Conservative 17; Mismatches 12; Indels 0; Gaps 0;

QY 1 MTNPPSAOIKPGEGFPLKIKARYDNFTGGWVAFADGCEYQNTPTVTGOLLCEVASSG 60
DB 35 MTNPPSSRIQPGEGFPLKIKARYDNFTGGWVADVDGGEYSNLTPTVTGOLCEIASSG 94
QY 61 KRDIIDLALDAAHKVKOKWAHTSVQDRAAIIKFKIADRMENLELLATAETWNGKPIRETS 120
DB 95 KRDIIDLALDAAHKAKDKWGHTSVQDRAAIIKFKIADRMENLELLATAETWNGKPIRETM 154
QY 121 AADVPLAIDHFRYFASCIRAQEGGISEVDSEIVAYHFHETGVVQIIPWNPFLMASWK 180
DB 155 AADVPLAIDHFRYFASCIRAQEGGISEVDKDIVAYHFHETGVVQIIPWNPFLMASWK 214
QY 181 MAPALAGNCVILKPARLTPLSVLLMEIVGDLPPGVNVVNGAGGVIGEYLATSKRIA 240
DB 215 MAPALAGNCVILKPARLTPLSVLLMEIVGDLPPGVNVVNGAGGEGEYLATSKRIA 274
QY 241 KVAFTGSTGVQGIIMQYATQNIIPVTLELGGKSPNIVADVMDDEDAFDDKALEGALFA 300
DB 275 KVAFTGSTGVQGIIMQYATQNIIPVTLELGGKSPNIVADVMDDEDAFDDKALEGALFA 344
QY 301 FNOGEVCTCPSRALVQESIVEREMERAIIRVESIRSGNPLDSVTOMGAUVSHGULETILN 460
DB 335 FNOGEVCTCPSRALVQESIVEREMERAIIRVESIRSGNPLDSVTOMGAUVSHGULETILN 494
QY 361 YIDIGKKEGADVLTGGRRKILLEGELKDGYYLEPTILFGONNRRVFOEIEFGVLAATTEK 420
DB 395 YIDIGKKEGADVLTGGRRKILLEGELKDGYYLEPTILFGONNRRVFOEIEFGVLAATTEK 454
QY 421 TMEALELANDTOYGLCAGVWSNPNLAYKMRGICQAGVWNTNCHAYPAHAAPGCKQS 480
DB 455 TMEALELANDTOYGLCAGVWSNPNLAYKMRGICQAGVWNTNCHAYPAHAAPGCKQS 514
QY 481 GIGRETHKMMLEHYQQTCKLLVSYSOKPLGLF 512
DB 515 GIGRETHKMMLEHYQQTCKLLVSYSOKPLGLF 546

RESULT 7
US-10-282-122A-72955
; Sequence 72955, Application US/10282122A
; GENERAL INFORMATION:
; APPLICANT: Wang, Lianqun
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: FLIIPA 934A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/220,948
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/207,727

```


; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PROP FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47174
 ; SEQ ID NO 15459
 ; LENGTH: 495
 ; TYPE: PRT
 ; ORGANISM: Xanthomonas campestris
 US-10-369-493-15459

Query Match 71.7% Score 1917; DB 6; Length 495;
 Best Local Similarity 71.9% Pred. No. 4.7e-167;
 Matches 353; Conservative 58; Mismatches 80; Indels 0; Gaps 0;

```

QY 22 KARYDNFICGEWAPADGEYONLTPTVTCQLCEVASSCKRDIIDLALDAAHKVKUKWAHT 81
DB 5 KPRYDNFICGEWAPADGEYONLTPTVTCQLCEVASSCKRDIIDLALDAAHKVKUKWAHT 64
QY 82 SVODRAAILFKIADRMFQNLLELLATAETWNGKPIRETSAAVDPLAIDHFRYFASCIRAQ 141
DB 65 STTDRSNVLKIAIDRIQNLLELLAYAEWTWNGKPVRETINADVPLCVDFHFRYFACAIRAQ 124
QY 142 EGGISEVDSETVAYHEHEPLGVVGQIIPWNPFLMASWMAKALAAAGNCVVLKPAKLTPL 201
DB 125 EGGISEIDSDTIAYHFHEPLGVVGQIIPWNPFLMACWKALAPALAAAGNCVVMKPAEQTPA 184
QY 202 SVLLMEIVGDLPPGVVNVVNGAGVIGEYIATSKRIAKVAFTGSTIEVGQIMOVATON 261
DB 185 SILVMEIVGDLPPGVVNVVNGFLEAGKPLASNPRIAKIAFTGETTIGRLIMOVASON 244
QY 262 IIPVTELGKSPNIFADVMDDEDAFFDKALGFALFNGEVCTCPSRALVOESIYE 321
DB 245 LIPVTELGKSPNIFADVMDDEDDFLDKAVEGFVLFAPNAGEVCTCPSRALIOESIYD 304
QY 322 RFMEKALRRVESIRSGNPLDSVTOMCAQVSHGQLEPILNYIDIGKEGADVLTGGRKLL 381
DB 305 TFMEKALRRVAALKQGNPLDPNTWGAQASSOLEKILSYIDIGKEGADVLTGGRKLL 364
QY 382 EGELKGYLLEPTILFGONNMRFQBEIFCPVLAVTTFTKMEALANDTOYGLGAGVW 441
DB 365 DGELSDGFYVKPITVFKGNKMRVQBEIFCPVSVTTTFKDEADALAIANDTIYGLGAGVW 424
QY 442 SRNCLAYKMGKGLQAGRVWNTNCYHAYPAHAAGGYKSGIGKRETHKMMLEHYQOTKLL 501
DB 425 SRASRLYRMGRLAIGRVWNTNCYHAYPAHAAGGYKSGIGKRETHKMMLEHYQOTKLL 484
QY 502 VSYSDKPLGLF 512
DB 485 VSYSPKALGFF 495
  
```

RESULT 10
 US-10-369-493-15827
 ; Sequence 15827, Application US/10369493
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PROP FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 15827
 ; LENGTH: 495
 ; TYPE: PRT
 ; ORGANISM: Xanthomonas campestris
 US-10-369-493-15827

Query Match 71.7% Score 1917; DB 6; Length 495;
 Best Local Similarity 71.9% Pred. No. 4.7e-167;
 Matches 353; Conservative 58; Mismatches 80; Indels 0; Gaps 0;

```

QY 22 KARYDNFICGEWAPADGEYONLTPTVTCQLCEVASSCKRDIIDLALDAAHKVKUKWAHT 81
DB 5 KPRYDNFICGEWAPADGEYONLTPTVTCQLCEVASSCKRDIIDLALDAAHKVKUKWAHT 64
QY 82 SVODRAAILFKIADRMFQNLLELLATAETWNGKPIRETSAAVDPLAIDHFRYFASCIRAQ 141
DB 65 STTDRSNVLKIAIDRIQNLLELLAYAEWTWNGKPVRETINADVPLCVDFHFRYFACAIRAQ 124
QY 142 EGGISEVDSETVAYHEHEPLGVVGQIIPWNPFLMASWMAKALAAAGNCVVLKPAKLTPL 201
DB 125 EGGISEIDSDTIAYHFHEPLGVVGQIIPWNPFLMACWKALAPALAAAGNCVVMKPAEQTPA 184
QY 202 SVLLMEIVGDLPPGVVNVVNGAGVIGEYIATSKRIAKVAFTGSTIEVGQIMOVATON 261
DB 185 SILVMEIVGDLPPGVVNVVNGFLEAGKPLASNPRIAKIAFTGETTIGRLIMOVASON 244
QY 262 IIPVTELGKSPNIFADVMDDEDAFFDKALGFALFNGEVCTCPSRALVOESIYE 321
DB 245 LIPVTELGKSPNIFADVMDDEDDFLDKAVEGFVLFAPNAGEVCTCPSRALIOESIYD 304
QY 322 RFMEKALRRVESIRSGNPLDSVTOMCAQVSHGQLEPILNYIDIGKEGADVLTGGRKLL 381
DB 305 TFMEKALRRVAALKQGNPLDPNTWGAQASSOLEKILSYIDIGKEGADVLTGGRKLL 364
QY 382 EGELKGYLLEPTILFGONNMRFQBEIFCPVLAVTTFTKMEALANDTOYGLGAGVW 441
DB 365 DGELSDGFYVKPITVFKGNKMRVQBEIFCPVSVTTTFKDEADALAIANDTIYGLGAGVW 424
QY 442 SRNCLAYKMGKGLQAGRVWNTNCYHAYPAHAAGGYKSGIGKRETHKMMLEHYQOTKLL 501
DB 425 SRASRLYRMGRLAIGRVWNTNCYHAYPAHAAGGYKSGIGKRETHKMMLEHYQOTKLL 484
QY 502 VSYSDKPLGLF 512
DB 485 VSYSPKALGFF 495
  
```

RESULT 11
 US-10-369-493-16210
 ; Sequence 16210, Application US/10369494
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PROP FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 16210
 ; LENGTH: 495
 ; TYPE: PRT
 ; ORGANISM: Xanthomonas campestris
 US-10-369-493-16210

Query Match 71.7% Score 1917; DB 6; Length 495;
 Best Local Similarity 71.9% Pred. No. 4.7e-167;
 Matches 353; Conservative 58; Mismatches 80; Indels 0; Gaps 0;

```

QY 22 KARYDNFICGEWAPADGEYONLTPTVTCQLCEVASSCKRDIIDLALDAAHKVKUKWAHT 81
DB 5 KPRYDNFICGEWAPADGEYONLTPTVTCQLCEVASSCKRDIIDLALDAAHKVKUKWAHT 64
  
```

QY 82 SWDRAALFKIADRMENLELLATAETWNGKPIKETSAAUPLAIDHPRFASQIRAQ 141
Db 65 STTDSNVLKIAIRIEQNLLELLAYAEFTWNGKPVPRFTINADVPLCVDHPRYPAGAIRAQ 124
QY 142 EGGISEVRSFTVAYHREPLGVVGGQIIPWNPPLIMASWKPAPALAAAGCNVVLKPPAPLTPDL 201
Db 125 EGGISEIDSDTIAYHREPLGVVGGQIIPWNPPLIMACWKLAPALAAAGCNVVKPAPQTPA 184
QY 202 SVLLIMEIVGDLPLPGVNVVNGAGCVIGEYLATSKRIAKVAFTGTSTEVGQQIMQVATON 261
Db 185 SILVIMEIVGDLPLPGVNVVNGFGLGAGKPLASNRIAKIAFTGTITGRIMQVASON 244
QY 262 IIPVTELEGGKSPNIVFADVDEDAFFDKALCEGALFAFNOCVECTCTSRALVORSIYE 321
Db 245 LIPVTELEGGKSPNIFADVMAELGAFIDRAVEGFLFAENQSEVCTCTSRALVQSINQ 304
QY 322 RMEHAIIRVRSIRSGNPLDSVTOMGACVSHGQLETLINVIDIGKREGADVLPGRPKLL 381
Db 305 TFMKALKRVAARIKGNPLDFTNMGACASSEQLFKLSYIDIGKJGJAEILLGGERNML 364
QY 382 EGEIKDGYILEPTILFGONNMRFQEEIFGVPVLAVTTFKTMEBALELANDTOYGLGAGVW 441
Db 365 DEGLSDGVPVFTFKHNRKRVFQEEIFGVPVSVTPKDEADALAIANDTLVGLGAGVW 424
QY 442 SRNGNLAYKMGRIQAGRVWTCYHAYPAHAAGGKYGKSGIGRETHKMMLEHYQOTKCLL 501
Db 425 SHDASRLVPMGRACIAGRVWTCYHAYPAHAAGGKYGKSGIGRETHKMMLEHYQOTKNI 484
QY 502 VSYSDKPLGLF 512
Db 485 VSYSPKALGFF 495

RESULT 12

US-10-282-122A-50457

: Sequence 50457, Application US/10282122A

: GENERAL INFORMATION:

: APPLICANT: Wang, Liangsu
: APPLICANT: Zamudio, Carlos
: APPLICANT: Malone, Cheryl
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ghisla, Kari
: APPLICANT: Zyskind, Judith
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John
: APPLICANT: Carr, Grant
: APPLICANT: Yamamoto, Robert
: APPLICANT: Forsyth, R.
: APPLICANT: Xu, H.

: TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

: FILE REFERENCE: ELITRA 034A

: CURRENT APPLICATION NUMBER: US/10/282,122A

: PRIOR FILING DATE: 2003-02-20

: PRIOR APPLICATION NUMBER: 60/191,078

: PRIOR FILING DATE: 2000-03-21

: PRIOR APPLICATION NUMBER: 60/206,848

: PRIOR FILING DATE: 2000-05-23

: PRIOR APPLICATION NUMBER: 60/207,727

: PRIOR FILING DATE: 2000-05-26

: PRIOR APPLICATION NUMBER: 60/230,335

: PRIOR FILING DATE: 2000-09-06

: PRIOR APPLICATION NUMBER: 60/230,347

: PRIOR FILING DATE: 2000-09-09

: PRIOR APPLICATION NUMBER: 60/242,578

: PRIOR FILING DATE: 2000-10-23

: PRIOR APPLICATION NUMBER: 60/254,625

: PRIOR FILING DATE: 2000-11-27

: PRIOR APPLICATION NUMBER: 60/257,931

: PRIOR FILING DATE: 2000-12-22

: PRIOR APPLICATION NUMBER: 60/267,636

: PRIOR FILING DATE: 2001-02-09

: PRIOR APPLICATION NUMBER: 60/269,308

: PRIOR FILING DATE: 2001-02-16

: Remaining Prior Application data removed - See File Wrapper or PALM.
: NUMBER OF SEQ. IN NOS: 78614
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 50457
: LENGTH: 567
: TYPE: PRT
: ORGANISM: Burkholderia mallei
US-10-282-122A-50457

Query Match

Best Local Similarity 71.6%; Score 1915; DB 6; Length 567;

Matches 357; Conservative 60; Mismatches 77; Indels 2; Gaps 1;

QY 17 FPLKLRKARYDNFIFGGEVWAPADGEYYQNLTPVTGOLLCEVASSCKRDLIDLADAAHKVD 76
Db 74 FP--YRQYGNFIFGGEVWAPVGGYFDNVSPVTRPPTAIPRSREADIELADAAHAARA 131
QY 77 KWAHTSVQDCAAIKFIADPMENQLELLATAFTWNGKPTFTSAAFVPIATIHPPYFAS 136
Db 132 GWAAGGAERANVLLPIADPMENITPLAVAFITDNCCKPLPETTAADVPLAIDHPPYPAG 191
QY 137 CTACACGCTSTSVDSFTVAYHREPLGVVGGQIIPWNPPLIMASWKPAPALAAAGCNVVLKPA 196
Db 192 CIRAQEGSIADIGGDMVAYHREPLGVVGGQIIPWNPPLIMAAWKLAPALAAAGCNVVLKPA 251
QY 147 PLTPLSVLLIMEIVGDLPLPGVNVVNGAGCVIGEYLATSKRIAKVAFTGTSTEVGQQIMQ 256
Db 252 EQTFASILVFAELIQDLPLPGVNVVNGFGLGAGKPLASNRIAKIAFTGTITGRIMQ 311
QY 257 YATONIIIPVTELEGGKSPNIVFADVDEDAFFDKALCEGALFAFNOCVECTCTSRALV 316
Db 312 YASENLIPVTELEGGKSPNIFADVMDPDDSYDFKALFGFAMFALNGCFVCTPSPALVE 371
QY 317 KSIYEREMERAIKPVFSTIRSCNPLDSVLIQMTACVSRHQLELILNYIDICKKCAUVLIG 376
Db 372 ESIYDFPERALKPVFAIKCHPLULSILMIGACASAPDQFKIUSYIDIGKJGJAEILLGG 431
QY 377 RKILIFGRKIKSGYYVLEPTILFGONNMRFQEEIFGVPVLAVTTFKTMEBALELANDTOY 436
Db 432 ERNVLGELAGBYVVKPTVFRGHKNKMRIFQEEIFGVPVLAVTTFKTEEEALELANDTL 491
QY 437 GAGVWSRNSNLAYKMGRIQAGRVWTCYHAYPAHAAGGKYGKSGIGRETHKMMLEHYQ 496
Db 492 GAGVWTRFNCNPAYRFFRGSTQAGRVWTCYHAYPAHAAGGKYGKSGIGRETHKMMLEHY 551
QY 497 TKCLLVSYSDKPLGLF 512
Db 552 TKNLLVSYSEKPLGFF 567

RESULT 13

US 10 369-493-12218

: Sequence 12218, Application US/10369493

: GENERAL INFORMATION:

: APPLICANT: Cao, Yongwei

: APPLICANT: Hinkle, Gregory J.

: APPLICANT: Slater, Steven C.

: APPLICANT: Goldman, Barry S.

: APPLICANT: Chen, Xianfeng

: TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

: TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

: FILE REFERENCE: 38-10(52953)B

: CURRENT APPLICATION NUMBER: US/10/369,493

: CURRENT FILING DATE: 2003-02-28

: PRIOR APPLICATION NUMBER: US 60/160,039

: PRIOR FILING DATE: 2002-02-21

: NUMBER OF SEQ. ID NOS: 47374

: SEQ ID NO 12218

: LENGTH: 505

: TYPE: PRT

: ORGANISM: Mesorhizobium loti

US-10-369-493-12218

```

1  : GENERAL INFORMATION:
2  : APPLICANT: Wang, Liansu
3  : APPLICANT: Zamudio, Carlos
4  : APPLICANT: Malone, Cheryl
5  : APPLICANT: Haselbeck, Robert
6  : APPLICANT: Ohlsen, Kari
7  : APPLICANT: Zyskind, Judith
8  : APPLICANT: Wali, Daniel
9  : APPLICANT: Trawick, John
10 : APPLICANT: Carr, Grant
11 : APPLICANT: Yamamoto, Robert
12 : APPLICANT: Forsyth, K.
13 : APPLICANT: Xu, H.
14 :
15 : TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
16 :
17 : FILE REFERENCE: ELTRA.034A
18 :
19 : CURRENT APPLICATION NUMBER: US/10/282,122A
20 : CURRENT FILING DATE: 2000-02-23
21 : PRIOR APPLICATION NUMBER: 60/191,078
22 : PRIOR FILING DATE: 2000-03-21
23 :
24 : PRIOR APPLICATION NUMBER: 60/206,848
25 : PRIOR FILING DATE: 2000-05-23
26 : PRIOR APPLICATION NUMBER: 60/207,727
27 : PRIOR FILING DATE: 2000-05-26
28 : PRIOR APPLICATION NUMBER: 60/230,335
29 : PRIOR FILING DATE: 2000-09-06
30 : PRIOR APPLICATION NUMBER: 60/240,347
31 : PRIOR FILING DATE: 2000-09-09
32 : PRIOR APPLICATION NUMBER: 60/242,578
33 : PRIOR FILING DATE: 2000-10-23

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:18:05, Search time 19.544 seconds
(without alignments)
2512.144 Million cell updates/sec

Title: us-09-830-751-8

Perfect score: 2673

Sequence: 1 MTNPPSAQIKPGYGFPLK... HYQTKCLVSYSDKPLGIF 512

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: PIR_73.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2668	99.8	542	1 S47809	aldehyde dehydrog
2	2651	99.2	542	2 H91186	aldehyde dehydrog
3	2651	99.2	542	3 C86033	aldehyde dehydrog
4	2589	96.9	512	2 AF0977	lactaldehyde dehyd
5	1918	71.8	506	2 R87693	aldehyde dehydrog
6	1910	71.5	506	1 A42597	aldehyde dehydrog
7	1895	70.9	505	2 AC2849	aldehyde dehydrog
8	1895	70.9	540	2 E97626	chloroacetaldehyde
9	1878	70.3	506	2 AF3735	aldehyde dehydrog
10	1849.5	69.2	515	2 H75589	aldehyde dehydrog
11	1848.5	69.2	505	2 AF3470	aldehyde dehydrog
12	1802	67.4	506	2 F83142	aldehyde dehydrog
13	1798	67.3	507	2 F70827	probable aldehyde
14	1791	67.0	506	2 H83396	probable aldehyde
15	1791	67.0	506	2 F82452	aldehyde dehydrog
16	1682	62.9	506	2 T09437	probable aldehyde
17	1682	62.9	541	2 A82276	aldehyde dehydrog
18	1085.5	40.6	495	2 C89778	hypothetical prote
19	1032	38.6	498	2 C83717	NADP-dependent ald
20	1010	37.8	501	2 T31265	aldehyde dehydrog
21	986	34.4	445	2 H66514	aldehyde dehydrog
22	981	36.7	519	1 I48966	aldehyde dehydrog
23	975	36.5	519	1 S03564	aldehyde dehydrog
24	971	36.3	517	1 DEHDE2	aldehyde dehydrog
25	970	36.3	503	2 T39216	aldehyde dehydrog
26	953	35.7	520	1 S09030	aldehyde dehydrog
27	952	35.6	500	1 S03364	aldehyde dehydrog
28	949	35.5	517	1 A40872	aldehyde dehydrog
29	940.5	35.2	519	1 S67286	probable aldehyde

30	940.5	35.2	520	1 S50576	probable aldehyde
31	932	34.9	544	2 D88442	protein F54D8.3 li
32	931.5	34.8	497	1 A29055	aldehyde dehydrog
33	930	34.8	542	2 T02301	aldehyde dehydrog
34	928	34.7	496	2 C87638	aldehyde dehydrog
35	925.5	34.6	499	2 S74224	aldehyde dehydrog
36	918.5	34.4	549	2 T03983	rf2 nuclear restor
37	918	34.3	538	2 I06583	aldehyde dehydrog
38	913.5	34.2	490	2 C83675	glycine betaine al
39	913	34.2	519	2 C86372	hypothetical prote
40	910	34.0	490	2 A64629	glycine betaine al
41	905	33.9	512	1 A55684	aldehyde dehydrog
42	895	33.5	509	1 S14629	aldehyde dehydrog
43	891	33.3	497	2 B82981	probable aldehyde
44	888.5	33.2	794	2 P95942	probable aldehyde
45	886.5	33.2	494	2 B95411	probable aldehyde

ALIGNMENTS

RESULT 1

S47809

aldehyde dehydrogenase (NAD) (EC 1.2.1.3) aldR [similarity] - Escherichia coli (strai
C:Species: Escherichia coli

C:Date: 10-Sep-1999 #sequence,revision 10-Sep-1999 #text_change 03-Jun-2002

C:Accession: S47809; B57259; F65158

R:Plunkett, G.

submitted to the EMBL Data Library, March 1994

A:Reference number: S47666

A:Accession: S47809

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-542 <PLU>

A:Cross-references: EMBL:U00032, NID:9466582, PDB:AAE1855, 1, PDB:9G12476

J. Xu, J. Johnson, R.C.

J. Bacteriol. 177, 3166-3175, 1995

A:Title: aldB, an RpoS dependent gene in Escherichia coli encoding an aldehyde dehydr

A:Reference number: A57259; MUID:95286498; PMID:7768815

A:Accession: B57259

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 31-542 <XUA>

A:Cross-references: GB:140742; NID:9712824; PDB:AAQ3693, 1, PDB:9712825

R. Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Petina, N.T.; Burkand, V.; Riley, M.,

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: F65158

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-542 <BLAT>

A:Cross-references: GB:AF000436, GB:U00096, NID:92367246, PDB:AAQ7662, 1, PDB:973900

A:Experimental source: strain F-12, substrain MG1655

C:Genetics.

A:Gene: aldB

A:Start codon: GTG

C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

C:Keywords: NAD, oxidoreductase

F:93.357/Domain: aldehyde dehydrogenase homology; ALDE

F:298.337/Active site: Glu, Cys #status predicted

F:494/Binding site: NAD (Cys) #status predicted

Query Match 99.8%, 50 to 2668, 78 1, length 542;

Best Local Similarity 99.8%; Pred No. 8.5e-198;

Matches 511, Conservative 0, Mismatches 1, Indels 0, Gaps 0;

QY 1 MTNPPSAQIKPGYGFPLKIKAHYUNFDSSEWVAALPKRYQUNLIPVIGQLIAEWASSIS 60

DB 31 MTNPPSAQIKPGYGFPLKIKAHYUNFDSSEWVAALPKRYQUNLIPVIGQLIAEWASSIS 90

QY 61 KPETCLALDAAKKVKKKAHNTSVQTFVAALFFKADHMEINLFLATAETWENKPIPIRETS 120

```

Db      91 KRDIIDLALDAHKVKWKWHTSVQDRAAILFKIADMEQNLELLATAETAETWNGKPIRETS 150
      121 AADVPLAIDHFRYFASCIQAEGGISEVDSVAYHFHEPLGVGVGQIIPWNPFLMLASWK 180
      151 AADVPLAIDHFRYFASCIQAEGGISEVDSVAYHFHEPLGVGVGQIIPWNPFLMLASWK 210
      181 MAPALAAGNCVVLKPARLTPLSVLLMEIVGDLPLPGVVNVVNGAGGVIAGEYLAISKRIA 240
      211 MAPALAAGNCVVLKPARLTPLSVLLMEIVGDLPLPGVVNVVNGAGGVIAGEYLAISKRIA 270
      241 KVAFTGSTEVGOQIMOVATONIPVTLELGGKSPNIFADVMDDEDAFFDKALEGFALFA 300
      271 KVAFTGSTEVGOQIMOVATONIPVTLELGGKSPNIFADVMDDEDAFFDKALEGFALFA 330
      301 FNOGEVCTCPSRALVQESIEYERFMEKAIKRVESIRSGNPIQAGRVWTCYHAYPAHAAGFGYKQS 480
      331 FNOGEVCTCPSRALVQESIEYERFMEKAIKRVESIRSGNPIQAGRVWTCYHAYPAHAAGFGYKQS 480
      361 YIDIGKKEGADVLTGGRKKLEGELEKDGYYLEPTILFGUNNNKRVQEEIFGPVLAATTFK 420
      391 YIDIGKKEGADVLTGGRKKLEGELEKDGYYLEPTILFGUNNNKRVQEEIFGPVLAATTFK 450
      421 TMEALELANDTOYGLGAGVWSRNGNLAYKMKRGICQAGRVWTCYHAYPAHAAGFGYKQS 480
      451 TMEALELANDTOYGLGAGVWSRNGNLAYKMKRGICQAGRVWTCYHAYPAHAAGFGYKQS 510
      481 GIGRETHKMMLEHYQOTKCLLYSYSDKPLGLF 512
      511 GIGRETHKMMLEHYQOTKCLLYSYSDKPLGLF 542

```

RESULT 2

```

H91186
aldehyde dehydrogenase B [imported] - Escherichia coli (strain O157:H7, substrain RMD 0
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug 2001
C:Accession: H91186
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kohara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and qenc
A:Reference number: A99620; MIMD:2115621; PMID:11258795
A:Accession: H91186
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-542 <HAY>
A:Cross-references: GB:BA000007; PIDN:BAH37887.1; PID:q13363935; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: Ecs4464
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

```

```

Query Match      99.2%; Score 2651; DB 2; Length 542;
Best Local Similarity 99.2%; Pred. No. 1.7e 196;
Matches 508; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy      1 MTNNPPSAQIKGEYGFPLKIKARYDNFICGSEWAPADGEYQNLTPVTGQLCEVASSG 60
      31 MTNNPPSAQIKGEYGFPLKIKARYDNFICGSEWAPADGEYQNLTPVTGQLCEVASSG 90
      61 KRDIIDLALDAHKVKWKWHTSVQDRAAILFKIADMEQNLELLATAETAETWNGKPIRETS 120
      91 KRDIIDLALDAHKVKWKWHTSVQDRAAILFKIADMEQNLELLATAETAETWNGKPIRETS 150
      121 AADVPLAIDHFRYFASCIQAEGGISEVDSVAYHFHEPLGVGVGQIIPWNPFLMLASWK 180
      151 AADVPLAIDHFRYFASCIQAEGGISEVDSVAYHFHEPLGVGVGQIIPWNPFLMLASWK 210
      181 MAPALAAGNCVVLKPARLTPLSVLLMEIVGDLPLPGVVNVVNGAGGVIAGEYLAISKRIA 240
      211 MAPALAAGNCVVLKPARLTPLSVLLMEIVGDLPLPGVVNVVNGAGGVIAGEYLAISKRIA 270

```

```

Qy      241 KVAFTGSTEVGOQIMOVATONIPVTLELGGKSPNIFADVMDDEDAFFDKALEGFALFA 400
      271 KVAFTGSTEVGOQIMOVATONIPVTLELGGKSPNIFADVMDDEDAFFDKALEGFALFA 440
      301 FNOGEVCTCPSRALVQESIEYERFMEKAIKRVESIRSGNPIQAGRVWTCYHAYPAHAAGFGYKQS 480
      331 FNOGEVCTCPSRALVQESIEYERFMEKAIKRVESIRSGNPIQAGRVWTCYHAYPAHAAGFGYKQS 480
      361 YIDIGKKEGADVLTGGRKKLEGELEKDGYYLEPTILFGUNNNKRVQEEIFGPVLAATTFK 420
      391 YIDIGKKEGADVLTGGRKKLEGELEKDGYYLEPTILFGUNNNKRVQEEIFGPVLAATTFK 450
      421 TMEALELANDTOYGLGAGVWSRNGNLAYKMKRGICQAGRVWTCYHAYPAHAAGFGYKQS 480
      451 TMEALELANDTOYGLGAGVWSRNGNLAYKMKRGICQAGRVWTCYHAYPAHAAGFGYKQS 510
      481 GIGRETHKMMLEHYQOTKCLLYSYSDKPLGLF 512
      511 GIGRETHKMMLEHYQOTKCLLYSYSDKPLGLF 542

```

RESULT 3

```

G86033
aldehyde dehydrogenase B [similarity] - Escherichia coli (strain O157:H7, substrain E
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Nov-2001
C:Accession: G86033
R:Perna, N.T.; Plunkett, III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MIMD:21074935; PMID:11206551
A:Accession: G86033
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-542 <SFO>
A:Cross-references: GB:AA005174; NID:q12518326; PIDN:AA058731.1; GSPDB:GN00145; UMDP:
A:Experimental source: strain O157:H7, substrain Ehu933
C:Genetics:
A:Gene: aldR
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

```

```

Query Match      99.2%; Score 2651; DB 2; Length 542;
Best Local Similarity 99.2%; Pred. No. 1.7e 196;
Matches 508; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy      1 MTNNPPSAQIKGEYGFPLKIKARYDNFICGSEWAPADGEYQNLTPVTGQLCEVASSG 60
      31 MTNNPPSAQIKGEYGFPLKIKARYDNFICGSEWAPADGEYQNLTPVTGQLCEVASSG 90
      61 KRDIIDLALDAHKVKWKWHTSVQDRAAILFKIADMEQNLELLATAETAETWNGKPIRETS 120
      91 KRDIIDLALDAHKVKWKWHTSVQDRAAILFKIADMEQNLELLATAETAETWNGKPIRETS 150
      121 AADVPLAIDHFRYFASCIQAEGGISEVDSVAYHFHEPLGVGVGQIIPWNPFLMLASWK 180
      151 AADVPLAIDHFRYFASCIQAEGGISEVDSVAYHFHEPLGVGVGQIIPWNPFLMLASWK 210
      181 MAPALAAGNCVVLKPARLTPLSVLLMEIVGDLPLPGVVNVVNGAGGVIAGEYLAISKRIA 240
      211 MAPALAAGNCVVLKPARLTPLSVLLMEIVGDLPLPGVVNVVNGAGGVIAGEYLAISKRIA 270
      241 KVAFTGSTEVGOQIMOVATONIPVTLELGGKSPNIFADVMDDEDAFFDKALEGFALFA 400
      271 KVAFTGSTEVGOQIMOVATONIPVTLELGGKSPNIFADVMDDEDAFFDKALEGFALFA 440
      301 FNOGEVCTCPSRALVQESIEYERFMEKAIKRVESIRSGNPIQAGRVWTCYHAYPAHAAGFGYKQS 480
      331 FNOGEVCTCPSRALVQESIEYERFMEKAIKRVESIRSGNPIQAGRVWTCYHAYPAHAAGFGYKQS 480
      361 YIDIGKKEGADVLTGGRKKLEGELEKDGYYLEPTILFGUNNNKRVQEEIFGPVLAATTFK 420
      391 YIDIGKKEGADVLTGGRKKLEGELEKDGYYLEPTILFGUNNNKRVQEEIFGPVLAATTFK 450

```

QY 421 TMEAEALANDTQYGLGAGVWSRNCNLAYKMGRIQAGRVWTCYHAYPAHAFAFGYKQS 480
 Db |||||
 QY 451 TMEAEALANDTQYGLGAGVWSRNCNLAYKMGRIQAGRVWTCYHAYPAHAFAFGYKQS 510
 Db |||||
 QY 481 GIGRETHKMMLEHYQOQTKCLLVSYSDKPLGLF 512
 Db |||||
 QY 511 GIGRETHKMMLEHYQOQTKCLLVSYSDKPLGLF 542
 Db |||||

RESULT 4

AF0977
 lactaldehyde dehydrogenase (EC 1.2.1.22) - Salmonella enterica subsp. enterica serovar Typhimurium
 C:Species: Salmonella enterica subsp. enterica serovar Typhimurium
 A:Note: this species has also been called Salmonella typhi
 C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AF0977
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Waig, J.; Churcher, C.; T. Connor, P.; Grodin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, J.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium
 A:Reference number: AB0502; PMID:11677608
 C:Accession: AF0977
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-512
 A:Cross-references: GH-AL51482; P1DN-CA00746.1; P1D-91504345; GSE68 GN00176
 C:Genetics: STY4116
 A:Gene: STY4116
 C:Superfamily: aldehyde dehydrogenase (NAD+) aldehyde dehydrogenase homology
 C:Keywords: oxidoreductase

Query Match 96.9%; Score 2589; DB 2; Length 512;
 Best Local Similarity 95.7%; Pred. No 97-192;
 Matches 490; Conservative 14; Mismatches 8; Indels 0; Gaps 0;

QY 1 MTNPPSAQIKPGEYGFLLKARYDNF IGGEWAPADAGEYQNLTPVTGQLCEVASSG 60
 Db |||||
 QY 1 MTNPPSAQIKPGEYGFLLKARYDNF IGGEWAPADAGEYQNLTPVTGQLCEVASSG 60
 Db |||||

QY 61 KRDI LALDAHKVKKWAHTSVQKAAALFKIARMEQNLLELATAETWNGKPIRETS 120
 Db |||||
 QY 61 KRDI LALDAHKVKKWAHTSVQKAAALFKIARMEQNLLELATAETWNGKPIRETS 120
 Db |||||

QY 121 AADVPLAIDHFRYFASCITRAQEGGISEVDSETVAYHFEPGLGVVGGQIIPWNPFLMASWK 180
 Db |||||
 QY 121 AADVPLAIDHFRYFASCITRAQEGGISEVDSETVAYHFEPGLGVVGGQIIPWNPFLMASWK 180
 Db |||||

QY 181 MAPALAAAGN'VVIKPARLTPILSVLLIMFTVGLIPPGVNVVNGAGVIGEYLATSKPIA 240
 Db |||||
 QY 181 MAPALAAAGN'VVIKPARLTPILSVLLIMFTVGLIPPGVNVVNGAGVIGEYLATSKPIA 240
 Db |||||

QY 241 KVAFTGSTFVGGQIMQYATQNIIPVTLGLGKSPNIVFADVMDDEDAFFDKALEGALFA 300
 Db |||||
 QY 241 KVAFTGSTFVGGQIMQYATQNIIPVTLGLGKSPNIVFADVMDDEDAFFDKALEGALFA 300
 Db |||||

QY 301 FNOGEVCT'PSPALVQESTYFPPFPPATPPVVESTPSGNPLDSVTOMGAQVSHGOLETILN 360
 Db |||||
 QY 301 FNOGEVCT'PSPALVQESTYFPPFPPATPPVVESTPSGNPLDSVTOMGAQVSHGOLETILN 360
 Db |||||

QY 361 YIDLUKKE'CAVAVLTHGKPKLLLEDELKUCYLYLEPTILFSGNNMKVFFEEIFSPVLAVTTFK 420
 Db |||||
 QY 361 YIDLUKKE'CAVAVLTHGKPKLLLEDELKUCYLYLEPTILFSGNNMKVFFEEIFSPVLAVTTFK 420
 Db |||||

QY 421 TMEAEALANDTQYGLGAGVWSRNCNLAYKMGRIQAGRVWTCYHAYPAHAFAFGYKQS 480
 Db |||||
 QY 421 TMEAEALANDTQYGLGAGVWSRNCNLAYKMGRIQAGRVWTCYHAYPAHAFAFGYKQS 480
 Db |||||

QY 481 GIGRETHKMMLEHYQOQTKCLLVSYSDKPLGLF 512
 Db |||||
 QY 481 GIGRETHKMMLEHYQOQTKCLLVSYSDKPLGLF 512
 Db |||||

RESULT 5

B87693
 aldehyde dehydrogenase B [imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
 C:Accession: B87693

P.Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.F.; Eisen, J.; Heidelberg, B.; Laub, M.T.; NeBry, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; H. H.; K. n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C. Proc Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.
 A:Reference number: A87249; PMID:21173698, PMID:11259647
 C:Accession: B87693
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-506
 A:Cross-references: GB AE005673; NID g13425422; P1DN AAK25542 1; GSPOR GN00148
 C:Genetics:

A:Gene: CC3580
 C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 71.8%; Score 1918; DB 2; Length 506;
 Best Local Similarity 71.7%; Pred. No. 5.5e-140;
 Matches 352; Conservative 57; Mismatches 82; Indels 0; Gaps 0;

QY 22 KARYDNF IGGEWAPADAGEYQNLTPVTGQLCEVASSGKKKLLALAAHKVKKWAHT 81
 Db |||||
 QY 16 KARYDNF IGGEWAPADAGEYQNLTPVTGQLCEVASSGKKKLLALAAHKVKKWAHT 75
 Db |||||

QY 82 SVUQPAALFKIARMEQNLLELATAETWNGKPIRETS AADVPLAIDHFRYFASCITRAQ 141
 Db |||||
 QY 76 SAADSPRILLPIADRMEEENLAALATAETWNGKPIRETS AADVPLAIDHFRYFASCITRAQ 135
 Db |||||

QY 142 EGGISEVDSETVAYHFEPGLGVVGGQIIPWNPFLMASWKMAPALAAAGN'CVLKPAPLTP 201
 Db |||||
 QY 136 EGGISEVDSETVAYHFEPGLGVVGGQIIPWNPFLMASWKMAPALAAAGN'CVLKPAPLTP 195
 Db |||||

QY 202 SVUQPAALFKIARMEQNLLELATAETWNGKPIRETS AADVPLAIDHFRYFASCITRAQ 261
 Db |||||
 QY 196 SIWVWAEWIGDILLPAGVNLVNGFGLGAGKPLASSPRIAKIATGHTSTGRLIMQYAAQN 255
 Db |||||

QY 252 IIPVTLGLGKSPNIVFADVMDDEDAFFDKALEGALFAFNOGEVCT'PSPALVQESTY 321
 Db |||||
 QY 256 IIPVTLGLGKSPNIVFADVMDDEDAFFDKALEGALFAFNOGEVCT'PSPALVQESTY 315
 Db |||||

QY 322 KMEPALPPVKSIFSGNPLDSVTOMGAQVSHGOLETILNIPVTLGLGKSPNIVFADVMDDEDAFFDKALEGALFA 381
 Db |||||
 QY 316 KMEPALPPVKSIFSGNPLDSVTOMGAQVSHGOLETILNIPVTLGLGKSPNIVFADVMDDEDAFFDKALEGALFA 375
 Db |||||

QY 382 EGGISEVDSETVAYHFEPGLGVVGGQIIPWNPFLMASWKMAPALAAAGN'CVLKPAPLTP 441
 Db |||||
 QY 376 PQGLADGYVEPTVFEHGNKMRIFQEEIFGVLAVTTFKTEEEALFIANDTAFGLGAGV 435
 Db |||||

QY 442 SPNGNLAYKMGRIQAGRVWTCYHAYPAHAFAFGYKQSGTGPTTHKMMI'PHYQOQTKCL 501
 Db |||||
 QY 436 SDANRCYRFGRIEAGRVWTCYHAYPAHAFAFGYKQSGTGPTTHKMMI'PHYQOQTKCL 495
 Db |||||

QY 502 VSYSDKPLGLF 512
 Db |||||
 QY 496 VSYSDKPLGLF 506
 Db |||||

RESULT 6

A42597
 aldehyde dehydrogenase (NAD) (EC 1.2.1.3) [validated] Alcaligenes eutrophus
 N:Alternate names: acetaldehyde dehydrogenase II
 C:Species: Alcaligenes eutrophus
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Jun-2002
 C:Accession: A42597

R:Priefert, H.; Kruger, N.; Jendrossek, D.; Schmidt, B.; Steinbuechel, A. J. Bacteriol. 174, 899-907, 1992

A:Title: Identification and molecular characterization of the gene coding for acetaldehyde
 A:Reference number: A42597; MUID:92121129; PMID:1732222

A:Accession: A42597

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-506 <PRI>

A:Cross-references: GR:M74003; NID:g141885; PIDD:AAA21943.1; PID:g141886

A:Experimental source: strain TF93

A:Note: sequence extracted from NCBI backbone (NCBIP:77618)

C:Genetics:

A:Gene: acd

C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

C:Keywords: NAD, oxidoreductase

F:57-321/Domain: aldehyde dehydrogenase homology <ALDD>

F:262-301/Active site: Glu, Cys #status predicted

F:456/Binding site: NAD (Cys) #status predicted

Query Match 71.5%; Score 1910; DB 1; Length 506;
 Best Local Similarity 71.5%; Pred. No. 2.3e-139;
 Matches 351; Conservative 67; Mismatches 73; Indels 0; Gaps 0;

QY 22 KARYDNFEGWVAPADGEYQNLTPVTGOLCEVASSCKRDLALDAHKKVKDKWAHT 81

DB 16 KQYENYIGAWPPAGGEYFESTTPTGKPTPRSQQDVAALDAHAAKAAWART 75

QY 82 SVODRAALFKIADRMENLELATAFTWDCNKPIRETSAAADVPLAIDHFRYFASCIRAQ 141

DB 76 STTERANILNRIADRIEANLKLVAESIDNCKPVRETAAADPLAVDHFYFACICIRAQ 135

QY 142 EGGISEVDSETVAYHFEPIGVVGGQIIPWNPFLIMASWKMAPALAAAGNCVVLKPARLTP 201

DB 136 EGGISEIDADTIAHFEPIGVVGGQIIPWNPFLIMATWKLAPALAAAGNCVVLKPAEQTPA 195

QY 202 SVLLMEIVGDLPPGVWVNVGAGVIGEXYATSKRIAKVAFSTGVGGQIMQYATON 261

DB 196 SILVMEIVGDLPPGVWVNVGAGVIGEXYATSKRIAKVAFSTGVGGQIMQYATON 255

QY 262 IIPVTELGKSPNIPVADVMEDEDAFFDKALEGAFALFAGVCTCPSPALVQESIYE 321

DB 256 LIPVTELGKSPNIPVADVMEDEDAFFDKALEGAFALFAGVCTCPSPALVQESIYE 315

QY 322 RMEHRAIRRVESIRSNPLDSTVOMGAOVSHQLETLINYYIDIGKKGADVLTCGRKLL 381

DB 316 RMEHRAIRRVESIRSNPLDSTVOMGAOVSHQLETLINYYIDIGKKGADVLTCGRKLL 375

QY 382 EGELKDGYYLEPTILFQNNMRVFOEIPCPVLAFTVTEKMEEALELANDTOYGLGAGW 441

DB 376 DGDLAGGYVVKPTVFAHGNKMRIFOEIPCPVLAFTVTEKMEEALELANDTOYGLGAGW 435

QY 442 SRNCLAYKMGRIQAGRWNTCYHAYPAHAAPGGYKSGIGRETHKMLHYYQOTKLL 501

DB 436 TRCARAFMRGRIQAGRWNTCYHAYPAHAAPGGYKSGIGRETHKMLHYYQOTKLL 495

QY 502 VSYSDKPLCLF 512

DB 496 VSYSPNALGPF 506

RESULT 7

AG2849

aldehyde dehydrogenase aldA [imported] - Agrobacterium tumefaciens (strain C58, Dupont)

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence revision 11-Jan-2002 #text change 01-Feb-2002

C:Accession: AG2849

R:Wood, D.W.; Sorubal, J.C.; Kaul, R.; Marks, P.; Chou, L.; Wood, G.E.; Chen, Y.; Woo, I.

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,

ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AG2577; PMID:11743193

A:Accession: AG2849

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-505 <KUR>

A:Cross-references: GR:AF008688; PIDD:AAI4213.1; PID:g1774065; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: aldA

A:Map position: circular chromosome

C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 70.9%; Score 1895; DB 2; Length 505;

Best Local Similarity 69.0%; Pred. No. 3.2e-138;

Matches 348; Conservative 68; Mismatches 84; Indels 4; Gaps 1;

QY 9 QIKPEGYGPEPLKAKYDNFICGEWVAPALGCHYYQNLTPVTGOLCEVASSCKRDLAL 68

DB 6 QKAKGEAPEKIK--YNYDGKWKPEKSGYMNISPVTDGKICFVPSNASHDEEAL 61

QY 69 IAAHKVKDKWARI SVGGRAAALIFKIAURMEONLELATAETWDCNKPIRETSAAALVPLAI 128

DB 62 DAAHKAKEKWKTSITERSNILLIKIAQRLEDNLLIARAETWDCNKPIRETSAAALVPLAI 121

QY 129 DHFRYFASCIAPAGEGGISEVDSETVAYHFEPIGVVGGQIIPWNPFLIMASWKMAPALAA 188

DB 122 DHFRYFAGCIRAQEGTIGEDNUTVAYHFEPIGVVGGQIIPWNPFLIMAAWKLAALAA 181

QY 189 NCVLPEAKLTPLSVLLMEIVGDLPPGVWVNVGAGVIGEXYATSKRIAKVAFSTGV 248

DB 182 NCVLPEAKLTPLSVLLMEIVGDLPPGVWVNVGAGVIGEXYATSKRIAKVAFSTGV 241

QY 249 EVGGQIMQYATONIPVTELGKSPNIPVADVMEDEDAFFDKALEGAFALFAGVCTCP 408

DB 242 SVGKEIMRYAANVTNISLELGKSPNIPVADVMEDEDAFFDKALEGAFALFAGVCT 401

QY 309 CPSKALVQESIYEFEPEFALIRVETSRNPLDSTVOMGAOVSHQLETLINYYIDIGKKE 468

DB 302 CPSPALVHVESIYHPEPEFALIRVETSRNPLDSTVOMGAOVSHQLETLINYYIDIGKKE 461

QY 369 GADVLIGSKRKLLEGELKIKYVLEPTILFQNNMRVFOEIPCPVLAFTVTEKMEEALE 428

DB 362 GAVLTGGDKTLTGDLKGGYIOPVPSKMKRIFOEIPCPVLAFTVTEKMEEALE 421

QY 429 ANDTQYGLGAGVWRNCLNAYKMGRIQAGRWNTCYHAYPAHAAPGGYKSGIGRETHK 488

DB 422 ANDTVYGLGAGVWRNCLNAYKMGRIQAGRWNTCYHAYPAHAAPGGYKSGIGRETHK 481

QY 489 MMLHYQOTKCLLVSYSDKPLCLF 512

DB 482 MMLDHTQOTKCLLVSYSPNKVGPF 505

RESULT 8

E97626

chloroacetaldehyde dehydrogenase (AF029733) [imported] - Agrobacterium tumefaciens (S

C:Species: Agrobacterium tumefaciens

C:Date: 30 Sep 2001 #sequence revision 30 Sep 2001 #text change 11 Jan 2002

C:Accession: E97626

R:Goodner, B.; Binkley, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, R.; Goldm

A.; Liu, P.; Woliam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz,

Science 294, 2323-2328, 2001

A:Title: Genome sequence of the plant pathogen and biotechnological agent Agrobacterium

A:Reference number: A97359; PMID:11743194

A:Accession: E97626

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-540 <KUR>

A:Cross-references: GR:AF007869; PIDD:AAK87966.1; PID:g1515174; GSPDB:GN00169

C:Genetics:

A:Gene: AGR_C_4041

A:Map position: circular chromosome

C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 70.9%; Score 1895; DB 2; Length 540;

Best Local Similarity 69.0%; Pred. No. 36e-138;
Matches 348, Conservative 68, Mismatches 84, Indels 4, Gaps 1;

QY 9 QIKRGYGFPLKAPYINFTGGHWAPAGCYEYQNIPTVTGQLLFEVASSSKKQIDIALDA 68
Db 41 QOKAGEAPFKLK---YGNVIGGKWVPEKSGRYMDNI SPVTGKHKEI EGVPSDASOTFFAL 96

QY 69 DAAHKVKDKWAHTSVQDPAATLTK (ADMPONI)FLATARTWNGKPIRETSAADVPLAI 128
Db 97 DAAHKAREKWKTSITERSNILLPTAQPTEDNLDLTPAETWNGKPLPETTNADIPLLI 156

QY 129 DHERYFASCTRAQEGGISEVDSETVAYHFHEPLGVVGGQITPWNFPLIMASWKMAPALAA 188
Db 157 DHERYFAGCIRAQEGGISEVDSETVAYHFHEPLGVVGGQITPWNFPLIMAAWKLAPALAA 216

QY 189 NCVVLKPARLTPSLVLLIMETVGDLLPPGVVNVVNGAGGVIGEVYGLATSKRIAKVFTGST 248
Db 217 NCVVLKPAEQTPASILVIMELIEDLLPTGVNLVINGTGLGAPLAGSRIAKIAFTGST 276

QY 249 EVGGQIMQATONITPVTLELGGKSNFIVFADYMDERDAFFUKALGGFALFAFNQGEVCT 308
Db 277 SVGKEIMEYAAICNVNTISLEIGSKSPNLEFALVNNETLAPLLKALGGFALFALNQEDEVCT 348

QY 309 GPSALVQPSIYFFHMPAPLPPVESIPSGNPLISVTQMTAGVASHGQLETTIINYIDGKKE 368
Db 337 GPSALVHESISYDFHMEKATKPVQALSQDEPLNPESTMAGASQDEPLKIMSYLEIGKKE 408

QY 369 GADVTGTPPKLLGGELKGGYLEPTILEGNNKPVFOEELFGPVLAVTTTEKTMERALEL 428
Db 397 GAKVLTGGDKTLTGDLKGGYIQPTVFEGNNKMKIFQEDIFGPVVSVTTEKKEVDALEI 456

QY 429 ANDTQYGGAGVWSRGNINAYKMGRTGQAGRVNTCYHAYPAHAAPGGYKQSGIGRETHK 488
Db 457 ANDTVYGLGAGVWSRDTNATYHAYPGVNTCYHAYVYAGAAFGYKQSGIGRETHK 516

QY 489 MWLEHYQVQKCLLVSYSDKPLGLF 512
Db 517 MWLDHYQVQKCLLVSYSPNKVGF 540

RESULT 9
A:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: AB3735
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: AB3660; MUID:20512582; PMID:11058132
A:Accession: AB3735
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-506 <STO>
A:Cross-references: GB:AP001509; GR:BA000004; NID:g10173176; PIRN:PAR04400.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 70.3%; Score 1878; DB 2; Length 506;
Best Local Similarity 68.9%; Pred. No. 6.7e-137;
Matches 346; Conservative 67; Mismatches 89; Indels 0; Gaps 0;

QY 11 KPGEYGFPLKARYDNFTGFWVAPADGEYQNIPTVTGQLLFEVASSSKKQIDIALDA 70
Db 5 RPNEGAVSFKKRYDNGVNGEWTTPVKGYFENVTPVTGEVCEVARSTAEIDIALDA 64

QY 71 AHKVKDKWAHTSVQDPAATLTK (ADMEQNL)ELLATAEIWNQKPIRETSAADVPLAIDH 130
Db 65 AHAKEQWKTSIPREPARNTLNKTADPMEENLEKLAETWENGKPVFTLNADIPLAIDH 124

QY 131 FRYFASCTRAQEGGISEVDSETVAYHFHEPLGVVGGQITPWNFPLIMASWKMAPALAA 190

Db 125 FRYFAGALRAVGGTSLSDNUTVAYHFHEPLGVVGGQITPWNFPLIMAWKLAALAAAGNC 184

QY 191 VWIPEAPLTPLSVLI LMFIVGDL PFGVNVVNVAGGVIGEVYGLATSKRIAKVFTGSTEV 250
Db 185 VILKPAEQTPASIFVLELIEDLLPKGVNINLVSPGVSEAGKPLASSPVAKVAFGTETTT 244

QY 251 QQQIMQATONITPVTLELGGKSNFIVFADYMDERDAFFUKALGGFALFAFNQGEVCTCP 310
Db 245 GRLTQYASENLIPVTLELGGKSPNIFDDVMKDDAFUNKALEGVLEALNQGVECTCP 304

QY 311 SPALVQESISYERFMEFAIPKVESIPSGNPLISVTQMTAGVASHGQLETTIINYIDGKKE 370
Db 305 SKALLIQESIVDTMERALARVQA LKQGNPLDPNTMICAQASQDEPLKIMSYLEIGKKE 364

QY 371 DVLTCGRKKLLELGGKGYLEPTILEGNNKPVFOEELFGPVLAVTTTEKTMERALELAN 430
Db 365 EVLAGGGRNHLFGELANGYVSPVTFKGTNDKMRVFOEELFGPVSVTTEKDAEEALATAN 424

QY 441 DVLTCGLGAGVWSRGNINAYKMGRTGQAGRVNTCYHAYPAHAAPGGYKQSGIGRETHKM 490
Db 445 DVLTCGLGAGVWSRGNINAYKMGRTGQAGRVNTCYHAYPAHAAPGGYKQSGIGRETHLMM 484

QY 491 LERHYQVQKCLLVSYSDKPLGLF 512
Db 495 LERHYQVQKCLLVSYSPNKVGF 506

RESULT 10
A:Species: Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: H75589
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
M., Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: H75589
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-515 <WHI>
A:Cross-references: GB:AE001863; GB:AE001825; NID:g6406070; PIRN:AAF12436.1; PID:g646
A:Experimental source: strain R1
C:Genetics:
A:Gene: DRA0348
A:Map position: 2
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 69.2%; Score 1849.5; DB 2; Length 515;
Best Local Similarity 69.1%; Pred. No. 1.1e-134;
Matches 346; Conservative 59; Mismatches 95; Indels 1; Gaps 1;

QY 12 KGEYGFPLKARYDNFTGFWVAPADGEYQNIPTVTGQLLFEVASSSKKQIDIALDA 71
Db 16 PGIFGVVTFKKRYDNEIFGGVWVFPVKGYFENVTPVTGEVCEVARSTAEIDIALDA 75

QY 72 HKVKDKWAHTSVQDPAATLTK (ADMEQNL)ELLATAEIWNQKPIRETSAADVPLAIDH 131
Db 76 HPAAPAWKPISVTHKSNILKIALPMEQNLLEMLAVAEIWNQKPIRETSAADVPLAIDH 135

QY 132 PYFASCTRAQEGGISEVDSETVAYHFHEPLGVVGGQITPWNFPLIMASWKMAPALAA 191
Db 136 PYFASCTRAQEGGISEVDSETVAYHFHEPLGVVGGQITPWNFPLIMASWKMAPALAA 195

QY 192 VLKPARLTPSLVLLIMETVGDLLPPGVVNVVNGAGGVIGEVYGLATSKRIAKVFTGSTEV 251
Db 196 VLKPAEQTPASIMVIMELIEDLLPGEVNVVNGAGGVIGEVYGLATSKRIAKVFTGSTEV 255

QY 252 QQIMQATONITPVTLELGGKSNFIVFADYMDERDAFFUKALGGFALFAFNQGEVCTCP 311
Db 256 QQIMQATONITPVTLELGGKSNFIVFADYMDERDAFFUKALGGFALFAFNQGEVCTCP 311

Db 256 RLIMGYAADNLIPVTLGLGKSPNIFDDVMEDUAFDLKAVGVMFALNGEVCCTCP 315

QY 312 RALVQESIYERFMRATRRVESIRSNPLDSVTOMGAQVSHQOETILNYIDIGKKEGAD 371

Db 316 RALVQESIYERFMRATRRVESIRSNPLDSVTOMGAQVSHQOETILNYIDIGKKEGAD 375

QY 372 VLTGGRKLLLEGKIDGYYLEPTILFGQNNMRFQEEIFGPVLAVTTFTKTEALELAND 431

Db 376 VLTGGRKLLLEGKIDGYYLEPTILFGQNNMRFQEEIFGPVLAVTTFTKTEALELAND 434

QY 432 TOYGLGAGVMSRGNLAYKMGRTQACRVWNTNCHYAPAHAAFGYKQSGIGRTHKMWL 491

Db 435 TLYGLGAGVMSRGNLAYKMGRTQACRVWNTNCHYAPAHAAFGYKQSGIGRTHKMWL 494

QY 492 EHYQOTKCLLVSYSDKPLGLF 512

Db 495 DHYQOTKCLLVSYSDKPLGLF 515

RESULT 11

AE3470

aldehyde dehydrogenase (NAD) (EC 1.2.1.3) [imported] - Brucella melitensis (strain 16M)

C:Species: Brucella melitensis

C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 03-May-2002

C:Accession: AE3470

R:DelVecchio, V.G.; Kaparral, V.; Pedkar, R.T.; Patra, G.; Mujor, C.; Los, T.; Ivanova, M.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessier, Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A:Title: The genome sequence of the facultative intracellular pathogen brucella melitensis

A:Reference number: AN3252; PMID:11756688

A:Accession: AE3470

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-505 <KUR>

A:Cross-references: GB:AE008417; PIRN:AA15528 1; PIR:gl7984776, GSPR:0000140

A:Experimental source: strain 16M

C:Genetics:

A:Gene: BME1747

A:Map position: 1

C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

C:Keywords: oxidoreductase

Query Match 69.2%; Score 1848.5; DB 2; Length 505;

Best Local Similarity 67.4%; Pred. No. 1,36-134;

Matches 341; Conservative 65; Mismatches 93; Indels 7; Gaps 1;

QY 7 SAOTKPGEGEPLKARYDNFEGGEWAPADREYQNLTPVTGLGCEVASSKKRDIOL 66

Db 7 SRVAKPA-----PAKRYCNFPGKWKVPPSPFPENTSPVNGGLVFPVAFSTAAVFA 59

QY 67 ALDAHKVKIKWAHTSVQDPAALPKIADPMQNLLELATAETWNGKPIRETSAAVPL 126

Db 60 ALDAHAAARELWGRTSVAERAILILNRADRIENLPALAAFTWNGKPIRETTNADLPL 119

QY 127 AIDHFRVFASCIQAQEGGISEVDSETVYHFEHPGLGVGQIIPWNPFLIMASWKPALA 186

Db 120 AVDFHFRFACVIRAQGGISFIDHDFVYHFEHPGLGVGQIIPWNPFLIMATWKLAPALA 179

QY 187 AGNCVVLKPARITPLISWLLIMFIVSPILPPGVVNVVNGAGVIGGLYATSKRIKAVFTG 246

Db 180 AGNCVVLKPARITPASILIMELIADIPPGVNVVNGGLEAGKPLASSPIAKIATG 239

QY 247 STEVQOQIMQYATONIIPTVLELGKSPNIVAIAMVDFEADAFUKALGSPALFAPNGCV 306

Db 240 ETTTGHITMGYASQNIIPVTLELGKSPNIFPKWAATFIDFIDKATPCPMFALNGGV 299

QY 307 CTCFSRALVQESIVEREMERATRRVESIRSGNPLDSVTOMGAQVSHQOETILNYIDIGK 366

Db 300 CTCFSRALVQESIVEREMERATRRVESIRSGNPLDSVTOMGAQVSHQOETILNYIDIGK 359

QY 367 KEGADVLTCGRKKLLLEGKIDGYYLEPTILFGQNNMRFQEEIFGPVLAVTTFTKTEALE 426

Db 360 QEGAELVAGNERMNLGGLAGGYVKKPTVFKGNKMKIKQEELFGPVVSVATFKDDAEAL 419

QY 427 ELANDTOYGLGAGVMSRGNLAYKMGRTQACRVWNTNCHYAPAHAAFGYKQSGIGRTH 486

Db 420 ELANDTOYGLGAGVMSRGNLAYKMGRTQACRVWNTNCHYAPAHAAFGYKQSGIGRTH 479

QY 487 HKMMLHYQOTKCLLVSYSDKPLGLF 512

Db 480 HKMMLHYQOTKCLLVSYSDKPLGLF 505

RESULT 12 F83142

probable aldehyde dehydrogenase PA4022 [imported] Pseudomonas aeruginosa (strain PA C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 41-Dec-2000

C:Accession: F83142

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lathig, R.J.L.; Lory, S.; Olson, M.V.

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A:Reference number: AB2950; PMID:2047337; PMID:10984043

A:Accession: F83142

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-506 <STO>

A:Cross-references: GB:AE004819; GB:AE004091; NID:g9950214; PIRN:AAG07409.1; GSPR:000

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA4022

C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 67.4%; Score 1802; DB 2; Length 506;

Best Local Similarity 67.1%; Pred. No. 4,96-141;

Matches 336; Conservative 59; Mismatches 106; Indels 0; Gaps 0;

QY 12 PGEYGFPLKARYDNFEGGEWAPADREYQNLTPVTGLGCEVASSKKRDIOLALAA 71

Db 6 PGEYGFPLKARYDNFEGGEWAPADREYQNLTPVTGLGCEVASSKKRDIOLALAA 65

QY 72 HKYDKKWAHTSVQDPAALPKIADPMQNLLELATAETWNGKPIRETSAAVPLADHP 141

Db 66 HAAAFAMWPTSVQDPAALPKIADPMQNLTPVTGLGCEVASSKKRDIOLALAA 126

QY 132 KTFASCIQAQEGGISEVDSETVYHFEHPGLGVGQIIPWNPFLIMASWKPALA 191

Db 126 KTFASCIQAQEGGISEVDSETVYHFEHPGLGVGQIIPWNPFLIMASWKPALA 185

QY 192 VLEPVLKPARITPLISWLLIMFIVSPILPPGVVNVVNGAGVIGGLYATSKRIKAVFTG 251

Db 186 VLEPVLKPARITPLISWLLIMFIVSPILPPGVVNVVNGAGVIGGLYATSKRIKAVFTG 245

QY 252 QOIMQYATONIIPTVLELGKSPNIVAIAMVDFEADAFUKALGSPALFAPNGCV 311

Db 246 QOIMQYATONIIPTVLELGKSPNIVAIAMVDFEADAFUKALGSPALFAPNGCV 305

QY 312 RALVQESIYERFMRATRRVESIRSNPLDSVTOMGAQVSHQOETILNYIDIGKKEGAD 371

Db 306 RALVQESIYERFMRATRRVESIRSNPLDSVTOMGAQVSHQOETILNYIDIGKKEGAD 365

QY 372 VLTGGRKLLLEGKIDGYYLEPTILFGQNNMRFQEEIFGPVLAVTTFTKTEALELAND 441

Db 366 VLTGGRKLLLEGKIDGYYLEPTILFGQNNMRFQEEIFGPVLAVTTFTKTEALELAND 425

QY 432 TOYGLGAGVMSRGNLAYKMGRTQACRVWNTNCHYAPAHAAFGYKQSGIGRTHKMWL 491

Db 426 TOYGLGAGVMSRGNLAYKMGRTQACRVWNTNCHYAPAHAAFGYKQSGIGRTHKMWL 485

QY 492 EHYQOTKCLLVSYSDKPLGLF 512

Db 486 EHYQOTKCLLVSYSDKPLGLF 506

RESULT 13

F70827

Probable aldehyde dehydrogenase (NAD) (EC 1.2.1.3) P70458 [similarity] - Mycobacterium tuberculosis

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 04-Jun-2002

C:Accession: F70827

R:Coile, S.T.; Brodeur, P.; Parkhill, J.; Garnier, T.; Churrier, C.; Harris, D.; Gordon, S.

R:Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

R:Jandream, M.A.; Rogers, J.; Fütterer, S.; Seeger, K.; Skellern, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: F70827

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-507 <COL>

A:Cross-references: db:AL021533, db:AL123456, NID:9346129, PIRN:AA17413.1, PIR:Q20954

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: RV0458

C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase homology: AL5-

C:Keywords: oxidoreductase

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

F:58-322/Domain: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

RESULT 14

H83396

Probable aldehyde dehydrogenase PA1984 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 16-Sep-2000 #sequence_revision 16-Sep-2000 #text_change 31-Dec-2000

C:Accession: H83396

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickney, M.J.; R

admat, S.; Yuen, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbia, K., L.

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

C:Reference number: A82350; MUID:2043747; PMID:1094043

A:Accession: H83396

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-506 <ST>

A:Cross-references: GR:AF004625, GR:AE004091, NID:99947983, PIRN:AAG55372.1, GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA1984

C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 67.08; Score 1791; DB 2; Length 506;

Best Local Similarity 66.74; Pred. No. 3 48-130;

Matches 332; Conservative 65; Mismatches 107; Indels 0; Gaps 0;

QY 12 PGYGFPLKLRKARYKINFGSEWVAFADGYYQNLTPVTGQLLCEVASSGKRDIDLALDA 71

Ib 5 PGIFSAVVIFKPPYGNVIGSEFPPVKSGYFINSVNGQPIAFPPPTAEIDIDALDA 65

QY 72 HKVKKWHTSVQDPAALFKIALRMENQLLELAIETWLNKIIFELSAALVPIADHF 131

Ib 66 HAAEAAGSPISVQEPSNLLKIALRLEQLLELAVETWLN:KAVPEILNADIPLAADHF 125

QY 132 PFASCIKALDEGGISEVSETVAYHPEHPLEPVVQIIPNFPILMASWMAALAAAGNCV 191

Ib 126 KYFAGCIRAGGSAAEINUSIVAYHIHEPLGVVQIIPNFPILMAAMKALALAAAGNCV 185

QY 192 VLKPARLTPLSVLLMEIVGDLPPGVVWVNVNAGSVIGEYLATSPTAKVAFSTGTEVG 251

Ib 186 VLKPARLTPLSVLLMEIVGDLPPGVVWVNVNAGSVIGEYLATSPTAKVAFSTGTEVG 245

QY 252 QVIMYALQNLIPVTLELGGKSPNIVFALVMEEDAPFALKALPGFALFAPNGEVCCTPS 311

Ib 246 SHILKCAENILIPSTVELGGKSPNIVFEDIMQAPAEIEKAAEGLVLAFFNGEVCCTPS 305

QY 312 PALVQESITPPFPERAIRRVESIRSNPLDSVTOMGAQVSHQLETLINVIIDIGKKEGA 371

Ib 306 PALVQESITPPFPERAIRRVESIRSNPLDSVTOMGAQVSHQLETLINVIIDIGKKEGA 365

QY 372 VLTGGPKLLEGLKDGYYLEPTILFGQNNMVFQEEIFGPVLAVTTFTEMEALELAN 431

Ib 366 LLAGSGVEKLEGNLASYIQTLLKGNMGVFOEIEFGVVGVTTFKDEAEALAIAND 425

QY 432 TQYGLGAGVWSRNCNLAAYKWSPIQACPVWNTCYHAYPAHAAPGGYKSGIGPETHKMM 491

Ib 426 TQYGLGAGVWSRNCNLAAYKWSPIQACPVWNTCYHAYPAHAAPGGYKSGIGPETHKMM 485

QY 492 EHYQQTICLLVSYSDKPLGLF 512

Ib 486 DHYQQTICLLVSYSDKPLGLF 506

RESULT 15

F82152

aldehyde dehydrogenase vcl1819 [imported] - Vibrio cholerae (strain N16961 serogroup O

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 02-Feb-2001

C:Accession: F82152

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Jackson, R.

R:Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers

L.; Park, M.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833; PMID:10952301

A:Accession: F82152

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-506 <HET>

A:Cross-references: GR:AE004259, GR:AE001852, NID:9956343, PIRN:AAF34967.1, GSPDB:GN

A: Experimental source: serogroup O1; strain N16961; biotype El Tor

C: Genetics:

A: Gene: VC1819

A: Map position: 1

C: Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

```

Query Match          67 0%; Score 1791; PR 2; Length 506;
Best Local Similarity 64.9%; Pred No. 3.4e-130;
Matches 326; Conservative 74; Mismatches 102; Indels 0; Gaps 0;

QY 11 KPGYGFPLKARKYDNFICGEWVAPADGEYQNLTPVTCQLICEVASSGKRDIDIALDA 70
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5 QPGSDNAVITFKSHYDNFICGQWVKPVSGEYFCGNISFVNGQVYCVQVAKSTQADIDALDA 64

QY 71 AHKYKDKWAHTSVQDRAALIFKTDADMEQNLELLATAETWDNCKPIRETSAADVPLADH 130
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
65 AHQVREAWAKTSVTERSNNLLKTDADRIEANIQLAVAEQWENGKPVRETAAADLPVVVDH 124

QY 131 FRYFASCIQAEGGISEVDSVETVAYHFHEPLGVVGQIIPNFPLLMASMKMAPALAAAGNC 190
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
125 FRYFAGCIRAQEGSAELDSHTASYHFPPEPGVVGQIIPNFPMLMAAKLAPALAAAGCC 184

QY 191 VVLKPARLTPLSVLLMEIVGDLDPGVVNVVNGAGGVIGEYLATSKRIAKVAFGTSTEV 250
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
185 VVLKPAEQTPTSILVLEIKIADLIPAGVLNVVNGFSGEAGQALATSQRIAKLAFTGSTQV 244

QY 251 GQIMQVATGNIIPVLEIAGKSPNIVFAVMEUEAFDFALEGFALFAFNGGEVCTCP 310
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
245 GQHILKCAOSLIPSTVEIGCKSPNIYFFDFDHEDTYLEKCIETGLIGFFNQGGEVCTCP 304

QY 311 SRALVQESIYERFMERAIRVESIRSGNPLDSVTOMGAQVSHGQLETIINYYIDICKKEGA 370
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
305 SRVLVHESIYDREFAKVAERAKGKOGNPLDTATQVGAQASQGFDPKILSYIDHIGRQEGA 364

QY 371 DVLTGGRKKLLEGELKNGYVLEPTILFGONNMHVFOEITFGPVIAVTTKTMEEALELAN 430
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
365 KVEGGVAKQEGELGGYTIQPTLLOGHNMKRVFOEITFGPVIAITSFKDEAEALAN 424

QY 431 DTQYGLGAGVMSRNCNLAAYKMGFGIQAGPVWTCNCHAYPAHAAPAGGYKQSGIGRETHKMM 490
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
425 DTEYGLGAGIWTWTDQNIAYHMGHNIQAGRIWNCNCHAYPAHAAPAGGYKQSGIGRETHKMM 484

QY 491 LEHYQQTCKLLVSYSDKPIGLF 512
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
485 LNHVONTKNLLISYDVNPIGLFF 506

```

Search completed: June 24, 2003, 10:18:33
Job time : 21.5932 secs


```

Query Match      99.8%   Score 2668   DB 1: Length 512;
Best Local Similarity 99.8%; Pred No 1 2 192;
Matches 511; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTNPPSAQIKPGEGYFPLKIKARYINFGKSWAPATGYYUNLTPVTGQLLCEVASSG 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 MTNPPSAQIKPGEGYFPLKIKARYINFGKSWAPADGYYUNLTPVTGQLLCEVASSG 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 61 KPDIDALADAAHKVKKWAHTSVQDPAALFKIADMEQNLLELLATAETWNGKPIRETS 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61 KPDIDALADAAHKVKKWAHTSVQDPAALFKIADMEQNLLELLATAETWNGKPIRETS 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 121 AADVPLAIDHFRYFASCIKRAOEGISEVDSEIVAVHFEPLGVGGQIIPWNPPLMASWK 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 121 AADVPLAIDHFRYFASCIKRAOEGISEVDSEIVAVHFEPLGVGGQIIPWNPPLMASWK 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 181 MAPALAAGNCVWLKPARLTPSLVLLMEIVGDLPPGVVNVVNGAGGIVGEYLATSKRIA 240
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 181 MAPALAAGNCVWLKPARLTPSLVLLMEIVGDLPPGVVNVVNGAGGIVGEYLATSKRIA 240
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 241 KVAFTGSTEVGGQIMQYATONITPVTLELGGKSPNIVFADVMDEEDAFDPALEGFALFA 300
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 241 KVAFTGSTEVGGQIMQYATONITPVTLELGGKSPNIVFADVMDEEDAFDPALEGFALFA 300
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 301 FNOGEVCTCPSRALVQESIVERPMEIRAIPRVESIRSGNPIDSVTOMGAQVSHGQLETLN 360
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 301 FNOGEVCTCPSRALVQESIVERPMEIRAIPRVESIRSGNPIDSVTOMGAQVSHGQLETLN 360
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 361 YIDIGKKEGAVLITGKRRKLELGEKIGYYLEPTILFGUNNNKRVFQELFGVPLAVTTFK 420
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 361 YIDIGKKEGAVLITGKRRKLELGEKIGYYLEPTILFGUNNNKRVFQELFGVPLAVTTFK 420
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 421 TMEALELANUTGYLGAGVWSKNGNLAYKMGKGLUAGRWVINCYNHAYPAHAAFGGYKQS 480
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 421 TMEALELANUTGYLGAGVWSKNGNLAYKMGKGLUAGRWVINCYNHAYPAHAAFGGYKQS 480
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 481 GIGRETHKMLLEHYQOTKCLLSYSKPLGLF 512
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 481 GIGRETHKMLLEHYQOTKCLLSYSKPLGLF 512
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 2
DH2A2_ALCEU0 STANDARD; PRT; 506 AA.
AC P46368;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Acetaldehyde dehydrogenase II (acoh) of Alcaligenes eutrophus.
GN ACOD.
OS Alcaligenes eutrophus (Ralstonia eutropha).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=510;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-20.
RX MEDLINE=92121129; PubMed 1732222;
RA Pfeiffer H, Krueger N, Jendrossek D, Schmidt B., Steinbuechel A.;
RT "Identification and molecular characterization of the gene coding for
RL J. Bacteriol 174:999-1007 (1992)
CC -1- FUNCTION: INVOLVED IN THE CATABOLISM OF ACETOIN AND ETHANOL.
CC -1- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O -> an acid + NADH.
CC -1- PATHWAY: Ethanol utilization, second step.
CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch)

```

```

CC FMPL M74093; AAA21943.1;
DR HSP; P05091; ICW3.
DR InterPro; IPR002086; Aldehyde_dehydr.
DR Pfam; PF06171; aldedh; 1.
DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD.
FT NTRIP 240 245 NAD (NUP PART) (BY SIMILARITY).
FT ACT_SITE 262 262 BY SIMILARITY.
FT ACT_SITE 301 301 BY SIMILARITY.
SQ SEQUENCE 506 AA; 54881 MW; A8715R094B126MD CR64;

Query Match      71.5%   Score 1910; DB 1: Length 506;
Best Local Similarity 71.5%; Pred. No. 8.7e-146;
Matches 351; Conservative 67; Mismatches 74; Indels 0; Gaps 0;

QY 22 KARYDNFPGGEVAPADGYYQNLTPVTGQLLCEVASSCKRIDLALDAAHKVKIKWAHT 81
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 16 KOQENVYICAWVPAGGEYFESTTPTCTKPTTRVPRSQQDVAALAAHAAKAAWKT 75
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 82 SVQDRAALIFK IADMEQNLLELLATAETWNGKPIRETSAADVPLAIDHFRYFASCIKRAQ 141
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 76 SITERANILNRIADRIEANKLIAVAESIDNGKPIRETTAAADPLAVIHFRYFAGCIKRAQ 145
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 142 EGGISEVDSEIVAVHFEPLGVGGQIIPWNPPLMASKMAIPALAAAGN'VVIKFAEGTFA 201
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 136 EGGISELDAIT IAYHFEPLGVGGQIIPWNPPLMATWKLALAAAGN'VVIKFAEGTFA 195
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 202 SVLLMEIVGDLPPGVVNVVNGAGGIVGEYLATSKRIAKVAFSGIHWGQIMQYATON 261
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 196 SVLLMEIVGDLPPGVVNVVNGAGGIVGEYLATSKRIASSKVAFTGTTTGTGIMQYATON 255
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 262 LIPVTLELGGKSPNIVFADVMDEEDAFDPALEGFALFAFNGEVC'VTSRALVQESIVE 421
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 256 LIPVTLELGGKSPNIFEDVLAADDAFFDPALEGFALFNGEVC'VTSRALVQESIVE 415
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 322 PMEPALIPRVESIRSGNPIDSVTOMGAQVSHGQLEPTILFGUNNNKRVFQELFGVPL 481
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 316 PMEPALIPRVESIRSGNPIDSVTOMGAQVSHGQLEPTILFGUNNNKRVFQELFGVPL 475
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 382 EGELKIGYYLEPTILFGUNNNKRVFQELFGVPLAVITTKTMEALELANUTGYLGAGVW 441
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 376 EGELKIGYYLEPTILFGUNNNKRVFQELFGVPLAVITTKTMEALELANUTGYLGAGVW 435
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 442 SRNGLAYKMGKGLUAGRWVINCYNHAYPAHAAFGGYKQSGIGRETHKMLLEHYQOTKCL 501
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 436 TRDCARAFRMGKGLUAGRWVINCYNHAYPAHAAFGGYKQSGIGRETHKMLLEHYQOTKCL 495
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 502 VSYSKPLGLF 512
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 496 VSYSPNALGFF 506
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3
DH2A2_DEIRA STANDARD; PRT; 515 AA.
AC Q9Y59; Q42502;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aldehyde dehydrogenase (EC 1.2.1.4).
GN ALDA OR DHA0448.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/deinococcus group; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI.
PX MEDLINE=2093696; PubMed=19567266;
PA White G., Fison J.A., Heideberg J.F., Hickey F.K., Petersen J.D.,
PA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,

```


RT dehydrogenase.";
 RL J. Bacteriol. 177:676-687(1995).
 CC -!- FUNCTION: DEGRADATES ALL ALDEHYDES POTENTIALLY GENERATED BY N
 CC DEALKYLATION OF THIOCARBAMATES AND MAY ALSO PARTICIPATE IN
 CC ETHANOLAMINE METABOLISM AND FURTHER ASSIMILATION OF DEGRADATION
 CC PRODUCTS BY THIOCARBAMATE-INDUCED CYTOCHROME P-450
 CC -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O -> an acid + NADH.
 CC -!- INDUCTION: BY EPTC (S-ETHYL DIPHOSPHYLAPRAMOTHYLALF)
 CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U17129; AAC77472.1; .
 DR HSSP: P05091; 1CW3.
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldehyd. 1.
 DR PROSITE: PS00670; ALDEHYDE_DEHYDR_CYS; 1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
 KW Oxidoreductase; NAD.
 FT INIT_MET 0
 FT NP_BIND 218 224 NAD (ADP PART) (BY SIMILARITY).
 FT ACT_SITE 262 262 BY SIMILARITY.
 FT ACT_SITE 301 301 BY SIMILARITY.
 FT ACT_SITE 302 302 BY SIMILARITY.
 SQ SEQUENCE 505 AA: 54863 MW: DBEDEC38B34B37 CRC64:

Query Match 66.2%; Score 1768.5; DB 1; Length 505;
 Best Local Similarity 65.9%; Pred. No. 3.6e-125;
 Matches 331; Conservative 68; Mismatches 102; Indels 1; Gaps 1.
 QY 11 KPGEYGFPLKLYKARYDNFICGEVAPADGEYQNIITPVTCGLLCEVASSCKRPDIALDA 70
 DB 5 KPGTADATMSQYDWNICNEWAPVKGQYFENIPVTGQNCPIWAKSTAEIDIELDA 64
 QY 71 ARKVKDKWHTSVQDRAAILFKIADRMENLIELATFWDNCKPIKETSAAIDVPLAIDH 130
 DB 65 AHAAPAWGKTSVAERAITLNTKATPMEKNLIELALAESWNCPIKPIETLNAIPLAIDH 124
 QY 131 PRYFASCIKAGGEGISEVDSEIVAYHPEHPIGVVQGIIPWFFPLMASWKMALAAAGN 140
 DB 125 FRYFAGATRAOEGSLSEINSDTVAYHPEHPIGVVQGIIPWFFPLMASWKMALAAAGN 184
 QY 191 VLKPARLTPLSVLLMEIVGDLPLPGVNVVNCAGGVIGEXLATSKRIAKVAFTGSTEV 250
 DB 185 IVLKPACTPVSILHLICIGDLLPAGVLNVNGFGEAGKPLASSPRIKKTAFTGETIT 244
 QY 251 GQIMQVATQNIIPVTLELGGKSNIVFADVMDDEDAFFDKALEGALFAPNNOGECTCP 310
 DB 245 GRLLMOVASQNLIPVTLELGGKSNIVFADVMDDEDAFFDKALEGALFAPNNOGECTAP 304
 QY 311 SRALVQESIEPRFPERAIRVESIRSNPLDSVYTCMAQVSHGSOLETIINYIDICKKEGA 370
 DB 305 SRALLQEDIFDEFLAMAAIRTKAVRQGDPIITDTMICAQNSDQLEKILSYIEIGKAEGA 364
 QY 371 DVLGTGRKRIIEGLIKMYYIEPTILFGQNNMVFQGEIIFGPGVLAVTTEKTEEAELELAN 430
 DB 365 KVIITGGPFAIGGPIISGYYVQPTVFTGNKNMIFQ-EIFGPPVSVTSFKDYDEALEIAN 423
 QY 431 DTQVGLGNGVSWSPNCLAYKMGEGIQAGVWNTWYHAYPAHAAPAGGYKAGGIPFTHKMM 440
 DB 424 ETLVGLGAGVMSREGGVAIFAGKULAGKRVMIINTYHQYPAHAAPAGGYKAGGIGRENHLM 443
 QY 491 LEHYQQTCKILVSYSDKPLGIF 512
 DB 484 LSHYQQTCKNLVSYAQAQAQGF 505

RESULT 7

DHAL_STRCO STANDARD; PPT: 507 AA
 ID DHAL_STRCO
 AC Q9R3Z6;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable aldehyde dehydrogenase (EC 1.2.1.3).
 GN SC01174 CH SC011A.05.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=2190410; PubMed=200953;
 FA Bentley S.D., Chater K.F., Gerdes-Barajas A.-M., Challis G.L.,
 FA Thomson N.P., James K.D., Harris D.E., Quail M.A., Kieser H.,
 FA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 FA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 FA Huang C.-H., Kieser T., Larke I., Murphy L., Oliver K., O'Neill S.,
 FA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
 FA Senger K., Saunders D., Sharp S., Squares P., Squares S., Taylor K.,
 FA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2)."
 EL Nature 417:141-147(2002).
 CC -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O -> an acid + NADH
 CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AL13210; CAB61586.1; .
 DR HSSP: P05091; 1CW3.
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldehyd. 1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
 DR PROSITE: PS00670; ALDEHYDE_DEHYDR_CYS; 1.
 KW Oxidoreductase; NAD; Complete proteome.
 FT NP_BIND 219 225 NAD (ADP PART) (BY SIMILARITY).
 FT ACT_SITE 263 263 BY SIMILARITY.
 FT ACT_SITE 302 302 BY SIMILARITY.
 SQ SEQUENCE 507 AA: 55673 MW: 0110815B2D8F96DD CRC64:
 Query Match 65.8%; Score 1760; DB 1; Length 507;
 Best Local Similarity 65.9%; Pred. No. 1.6e-124;
 Matches 330; Conservative 64; Mismatches 107; Indels 0; Gaps 0;
 QY 12 PGEYGFPLKLYKARYDNFICGEVAPADGEYQNIITPVTCGLLCEVASSCKRPDIALDA 71
 DB 7 PGTETALIVSYSPYKHPIEGEYVPPAPGQYFFENSPVNIIPHTETAGTADIVPALDA 66
 QY 72 HKVVDKWHSTSVQDRAAILFKIADRMENLIELATFWDNCKPIKETSAAIDVPLAIDH 131
 DB 67 HEATPGVQPTSVTEPSIILKINPMPANI.EPI.AVAFSWPNCPIVETLANIPLAIDH 126
 QY 131 PRYFASCIKAGGEGISEVDSEIVAYHPEHPIGVVQGIIPWFFPLMASWKMALAAAGN 191
 DB 127 FRYFAGATRAOEGSLSEINSDTVAYHPEHPIGVVQGIIPWFFPLMASWKMALAAAGN 186
 QY 191 VLKPARLTPLSVLLMEIVGDLPLPGVNVVNCAGGVIGEXLATSKRIAKVAFTGSTEV 251
 DB 187 VLKPARLTPLSVLLMEIVGDLPLPGVNVVNCAGGVIGEXLATSKRIAKVAFTGSTEV 246
 QY 251 GQIMQVATQNIIPVTLELGGKSNIVFADVMDDEDAFFDKALEGALFAPNNOGECTCP 311
 DB 245 GRLLMOVASQNLIPVTLELGGKSNIVFADVMDDEDAFFDKALEGALFAPNNOGECTAP 311

DB 247 RLIMQYASENIKPVTLELOCKSPNIFEDVWARDDDFRKALEGTFMFPALNOGEVCTQPS 306
 QY 312 RALVQESIYERFERAIRRVERSTRSGNPLISVTCMAQVSHQJLETLINYLIDIGKEGAD 371
 DB 307 RALVORGVYAEFEAAVARETELKPGHPLDPTDMICQAQSNPQLEKILSYLDIGROGAK 366
 QY 372 VLTGGRKILLEGELKDCGYLEPTILFGQNNMRVFEIEFGPVLAVTTFTKMEALELAND 431
 DB 367 VLTGGERLEHDEGLKGYGVQPTIFGHNPMRIFQEEIEFGPVSVTSFDDFDVAKVTAND 426
 QY 432 TOYGLGAGVWSRNGNLAYKMGRIQAGRVWNTNCYHAYFAHAFAGGYKQSGIGRETHKMWL 491
 DB 427 TLYGLGAGVWNTDMNTAYRAGRAIQAGRVWNTNCYHAYFAHAFAGGYKQSGIGRENKMWL 486
 QY 492 EHYQOTKCLLVSYSDKPLGLF 512
 DB 487 EHYQOTRNLICSYSPKRLGFF 507
 RESULT 8
 ID DHAL_VIBCH STANDARD; PRT; 506 AA.
 AC P23240: Q9KTS0;
 DT 01-NOV-1991 (rel. 20, Created)
 DT 01-NOV-1991 (rel. 20, last sequence update)
 DT 15-JUN-2002 (rel. 41, last annotation update)
 DE Aldehyde dehydrogenase (EC 1.2.1.3).
 GN ALDA OR VC0819.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
 OX NCHI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KPR 56.
 RX MEDLINE=91210174; PubMed=1902210;
 RA Parsot C., Mekalanos J.J.;
 RT "Expression of the Vibrio cholerae gene encoding aldehyde
 RT dehydrogenase is under control of ToxR, the cholera toxin
 RT transcriptional activator.";
 RL J. Bacteriol. 173:2842-2851(1991).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=El Tor N16961 / Serotype O1;
 RX MEDLINE=98169509; PubMed=9501228;
 RA Karaolis D.K.R., Johnson J.A., Bailey C.C., Boedeker E.C., Kaper J.B.,
 RA Reeves P.R.;
 RT "A Vibrio cholerae pathogenicity island associated with epidemic and
 RT pandemic strains.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:3134-3139(1998).
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Classical Ogawa 395 / ATCC 39541 / Serotype O1;
 RX MEDLINE=21117025; PubMed=11179381;
 RA Karaolis D.K.R., Ian R., Kaper J.B., Reeves P.R.;
 RT "Comparison of Vibrio cholerae pathogenicity islands in sixth and
 RT seventh pandemic strains.";
 RL Infect. Immun. 69:1947-1952(2001).
 [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=El Tor N16961 / Serotype O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Emolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 RT cholerae.";
 RL Nature 406:477-483(2000).
 CC -1: FUNCTION: MAY BE INVOLVED IN V. CHOLERAEE VIRULENCE, AS ITS
 CC EXPRESSION IS UNDER THE CONTROL OF TOXR, A TRANSCRIPTIONAL

CC ACTIVATOR OF SEVERAL GENES ASSOCIATED WITH VIRULENCE.
 CC -1: CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O -> an acid + NADH.
 CC -1: PATHWAY: Ethanol utilization; second step.
 CC -1: SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb.sib.ch/](http://www.isb.sib.ch/announcement/)
 CC or send an email to license@isb.sib.ch).
 CC -----
 DB EMBL: M60658; AAA03051.1;
 DB EMBL: AF034434; AAC1274.1;
 DB EMBL: AF325733; AAK20747.1;
 DB EMBL: AE004167; AAF93982.1; ALT_INIT.
 DB HSSP: P05091; 1CW3.
 DB TIGR: VC0819;
 DB InterPro: IPR002086; Aldehyde_dehydr.
 DB Pfam: PF00171; aldedh_1.
 DB PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
 DB PROSITE: PS00687; ALDEHYDE_DEHYDR_GLD; 1.
 KW Oxidoreductase; NAD; Complete proteome.
 FT NP_BI: 240 245 NAB (ADP PART) (BY SIMILARITY).
 FT ACT_SITE: 262 263
 FT ACT_SITE: 301 301 BY SIMILARITY.
 FT ACT_SITE: 401 401 BY SIMILARITY.
 SQ SEQUENCE 506 AA; 55884 MW; 69E7F561D6CFE07 CR64;
 Query Match 62.9%; Score 1682; DB 1; Length 506;
 Best local similarity 63.4%; Prod. No. 1.1e-118;
 Matches 309; Conservative 67; Mismatches 112; Indels 0; Gaps 0;
 QY 25 YDNFEGSEWVAPADGEGYUNLTHVTVQLLEEVASSKKRDLALALAAUKUKIAHISVQ 84
 DB 19 YDNFEGSEWVAPADGEGYUNLTHVTVQLLEEVASSKKRDLALALAAUKUKIAHISVQ 78
 QY 85 DRAAILFKIADRMEONLELTAETWNGKPIRETSAADVPLAIDHRYFASCIQAQPG 144
 DB 79 ERSNILLRIADRIESNLETLAIVESMNGKPIRETIAALPLIIDHRYFAANIKSOGCA 138
 QY 145 ISEVDSFTVAYHPEHPEGVVGGQIIPNFPPLIMASKMAIPLAANGVVLKARITPESVL 204
 DB 139 ASLEDSKTIITYHLPEDPGVGGQIIPNFPPLIMAAWRIALPLAANGVVLKARITPESVL 198
 QY 205 LIMEIVGDLPLPGVWVWVNVNAGVIGELYLATSKRIAKVAFGTGTEWGGQIDQYATONIIP 264
 DB 199 FIMEIIGDLIPAGVINNVNFGSEAGNALATSGRIDKLAFTGTSTEIGNHILKCAADNLIIP 258
 QY 265 VTLEIGCKSPNIVFAIVMDEHDAFFIKALEGALFAPNCGVCTCTDSALVQESIERFM 324
 DB 255 STLEIGCKSPNIVFAIVMDEHDAFFIKALEGALFAPNCGVCTCTDSALVQESIERFM 318
 QY 325 FPAIPVSESIPSNIPDSVTCMAQVSHQJLETLINYLIDIGKEGAD 384
 DB 319 AKILIERVALIKQNPPLDTEIQCAQVSKQYDIKILCYIQIKDQDAELIFGHPNNQENY 478
 QY 385 LKDGYYLEPTILFGQNNMRVFEIEFGPVLAVTTFTKMEALELANDYQYGLAGVWSRN 444
 DB 379 LSGGYVVKPTLFGHNMHIFQEEIEFGPVVIAITKFKDEIEALHLANITVYGLGAGVWTRD 438
 QY 445 GNLAYKMGRIQAGRVWNTNCYHAYFAHAFAGGYKQSGIGRETHKMWLEHYQOTKCLVSY 504
 DB 439 INTAHPMAKNIKAGRPVWVWVNVNAGVIGELYLATSKRIAKVAFGTGTEWGGQIDQYATONIIP 504
 QY 505 SUKPLGLF 512
 DB 499 EHPHGLGF 506
 RESULT 9
 DHAM_MOUSE
 ID DHAM_MOUSE STANDARD; PRT; 519 AA.

RL Arch Biochem Biophys. 289:214-223(1991).
 CC -|- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
 CC -|- PATHWAY: Ethanol utilization; second step.
 CC -|- SUBUNIT: HOMOTETRAMER.
 CC -|- CELLULAR LOCATION: Mitochondrial matrix.
 CC -|- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X14977; CAA33101.1; -;
 DR EMBL; M19030; AAA40719.1; -;
 DR PIR; A27113; A27113;
 DR PIR; S03564; S03564;
 DR PIR; S17492; S17492;
 DR HSSP; P05091; ICW3;
 DR InterPro; IPR002086; Aldehyde_dehydr.
 DR Pfam; PF00171; aldedh; 1;
 DR PROSITE; PS00070; ALDEHYDE_DEHYDR_GYS; 1;
 DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1;
 KW Oxidoreductase; NAD; Acetylation; Mitochondrion; Transit peptide.
 FT TRANSIT 1 19 MITOCHONDRION.
 FT CHAIN 20 519 ALDEHYDE DEHYDROGENASE.
 FT MOD_RES 20 20 ACETYLATION (PROBABLE).
 FT NP_BIND 264 269 NAD (ADP PART) (BY SIMILARITY)
 FT ACT_SITE 287 287 BY SIMILARITY.
 FT ACT_SITE 321 321 BY SIMILARITY.
 SQ SEQUENCE 519 AA; 56488 MW; 75C748202F1333E5 CRC64;

 Query Match 36.5%; Score 975; DB 1; Length 519;
 Best Local Similarity 42.4%; Pred. No. 1.2e-65;
 Matches 208; Conservative 86; Mismatches 172; Indels 24; Gaps 10.

 QY 28 FIGEWAPADGVEYVNIPTVTGCLCEVASSGKRDIDLAL---DAAHKVKDKWAHTSVQ 84
 DB 42 FINNEHDVASKKTPFTVNFSTGEVICVAEGNKEDVDKAVKAAQAALQSGPWRMDAS 101

 QY 85 DRAALFKADREQNLELLATETWNGKPIRETSAAVPLAIDHFRFASCIRAQEGG 144
 DB 102 DRGLVRLADLERDRTYLAALLETLDNGKPYVVISYLVLDLWLVLCRLRYAGWADKYHCK 161
 QY 145 ISEVDSITVAYHFEHPLGVVGQIIPWNPFLMASNMKMAPALAAAGNCVILKPARLTPLSVL 204
 DB 162 TTPIDGFFSYTHREPVGVGQIIPWNPFLMLMANKLGFALAGNVVWVKVAEQIPLTAL 221

 QY 205 LLMPTVQL-IPPSVNVVNGAGSVIGEYLATSKKIAKVAFITGSTEVSQGIQMYA-TUNI 262
 DB 222 VVANLKEAGFPFCVNVIVGFTGTAGAAIASLHSDVDKVAFTGSTEVGHLLIQVAAGSSNL 281

 QY 263 IPTVLEGGKSPNIVADWDEDAFDFKALE--GFALFAPNQGEVCTCPSPALVQESLY 320
 DB 282 KRVTLLEGKSPNIVMS-----DAIMFWAVFQAIPALF-FNQGQCCAGSFTEFQEDVY 334

 QY 321 ERMERAIRVESIRSGNPLDSVTOMGAQVSHQLETTILNYIDIGKKEGADVLTGGRKL 380
 DB 335 DEEVSERVAAPAKSRVVCNFTSPTETSPVETPKILGYIKSQAQEAQKLLDGG---- 390

 QY 381 LEGELKD-GYILEPTILFG--QNNMVKVFOFEIFGPIVLAFTPKTMEHALELANDTQYGLG 437
 DB 391 --GAADRGVFIQPTV-FGIVDKGMTIAKFEITFGVPMQILKPKTTFEEVVGRRANSKYGIA 447

 QY 438 AGVMSRNGLAYKMGKGTGACGAPWTNCTYHAYPAHAFAFGYKSGKSGIGRETHKMMLIHYOOT 497
 DB 448 AAVFTKULDKANVLSALQAGTWTINCYIVFCAQSPFGYKMSGSCRELFGVFGIQAFTFV 507

 QY 498 KCLLVSYSDK 507
 DB 508 KTVTVKVPQK 517

RESULT 11
 DHAM_HUMAN
 ID DHAM_HUMAN STANDARD; PRT; 517 AA.
 AC P05091; Q03639;
 DT 13-AUG-1987 (Rel. 05, Created)
 DI 01-JAN-1990 (Rel. 13, Last sequence update)
 FT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Aldehyde dehydrogenase, mitochondrial precursor (E.C. 1.2.1.3) (ALDH
 DE class 2) (ALDH2) (ALDH-E2).
 GN ALDH2 OR ALDH.
 GN Homo Sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCHL_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE 88256152; PubMed 2848413;
 RA Hsu L.C., Hendel R.E., Yoshida A.;
 RT "Genomic structure of the human mitochondrial aldehyde dehydrogenase
 RT gene.";
 RL Genomics 2:57-65(1988).
 RN [2]
 RP SEQUENCE OF 18-517.
 RC TISSUE=Liver;
 RX MEDLINE 86055846; PubMed 4065146;
 RA Hempel J., Kaiser R., Joernvall H.;
 RT "Mitochondrial aldehyde dehydrogenase from human liver. Primary
 RT structure, differences in relation to the cytosolic enzyme, and
 RT functional correlations.";
 RL Eur. J. Biochem. 153:13-28(1985).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Muscle;
 RX MEDLINE 87174836; PubMed 3562250;
 RA Braun T., Bober E., Singh S., Agarwal D.P., Goedde H.W.;
 RT "Isolation and sequence analysis of a full length cDNA clone coding
 RT for human mitochondrial aldehyde dehydrogenase.";
 RL Nucleic Acids Res. 15:3179-3179(1987).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Muscle;
 RX MEDLINE 87219091; PubMed 3582651;
 RA Braun T., Bober E., Singh S., Agarwal D.P., Goedde H.W.;
 RT "Evidence for a signal peptide at the amino terminal end of human
 RT mitochondrial aldehyde dehydrogenase.";
 RL FEBS Lett. 215:233-236(1987).
 RN [5]
 RP REVISIONS TO N-TERMINUS.
 RA Braun T., Bober E., Singh S., Agarwal D.P., Goedde H.W.;
 RL FEBS Lett. 233:440-440(1988).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RA Strausberg K.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBS databases.
 RN [7]
 RP SEQUENCE OF 119-517 FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE 85216574; PubMed 2987944;
 RA Hsu L.C., Tani K., Fujiyoshi T., Kurachi K., Yoshida A.;
 RT "Cloning of cDNAs for human aldehyde dehydrogenases 1 and 2.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4771-4775(1985).
 RN [8]
 RP SEQUENCE OF 119-517 FROM N.A.
 RX MEDLINE 85252089; PubMed 4015824;
 RA Yoshida A., Ikawa M., Hsu L.C., Tani K.;
 RT "Molecular abnormality and cDNA cloning of human aldehyde
 RT dehydrogenases.";
 RL Alcohol 2:104-106(1985).
 RN [9]
 RP SEQUENCE OF 214-500 FROM N.A.

Db 446 AAVFTKDLDRANYSLSQALQAGTVMNCLYDFGCAUSFFGGYKMSGSGRELGEGLQAYTEV 505
 QY 498 KCLLVSYSDK 507
 Db 506 KTVTKVPQK 515

RESULT 12

DHAM_MESAU ID DHAM_MESAU STANDARD; PPT; 500 AA.
 AC P81178;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Aldehyde dehydrogenase, mitochondrial (EC 1.2.1.3) (ALDH class 2)
 DE (ALDH1) (ALDH-E2).
 GN ALDH2.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria, Rodentia, Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus
 OX NCBI_TaxID=10036;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Liver;
 PX MEDLINE=98034175; PubMed=9369242;
 RA Hjelqvist L., Lundgren R., Norin A., Joernvall H., Vallee B., Klyosov A., Keung W.M.;
 RT "Class 2 aldehyde dehydrogenase. Characterization of the hamster enzyme, sensitive to daldin and conserved within the family of multiple forms.";
 RT FEBS Lett. 416:99-102(1997).
 CC -|- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
 CC -|- PATHWAY: Ethanol utilization; second step.
 CC -|- SUBUNIT: HOMOTETRAMER.
 CC -|- SUBCELLULAR LOCATION: Mitochondrial matrix.
 CC -|- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 DR HSSP: P05091; 1CW3.
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldedh; 1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GH; 1.
 DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
 KW Oxidoreductase; NAD; Mitochondrion.
 FT NP_BIND 245 250 NAD (ADP PART) (BY SIMILARITY).
 FT ACT_SITE 268 268 BY SIMILARITY.
 FT ACT_SITE 302 302 BY SIMILARITY.
 SQ SEQUENCE 500 AA; 54334 MW; 8B81866AB04F493F CRC64;

Query Match 36.2%; Score 968; DR 1; Length 500;

Best Local Similarity 42.2%; Pred. No. 3 40-65;

Matches 207; Conservative 86; Mismatches 173; Indels 24; Gaps 10;

QY 28 FIGEWVAPADGVEYQNIPTVTGQLCEVASSGKRDIDLALDAH---KVKDKWAHTSVQ 84
 Db 23 FINNEWDHVSCKTPTPTNPSTGEVICQVAESKEDVIKAVKAAAFQLGSPWPRMDAS 82
 QY 85 DRAILFKIADRMQLELLATAETWNGKPIRETSAADVPLAIDHFRYFASCIQAEGG 144
 Db 83 DRGRLNRLADLIERDRTYLAALTLNCKPVYISVLDLMDVLCRLRYAGWADKYHCK 142
 QY 145 ISEVDSHTVAYHPEPLGVGVGOLIPNPFLLMASMKMAPALAGNCVVKARITPLSVL 204
 Db 143 TPIPTGQFTSTHFFPVWVGGTIPWNEFLLWAKLGHALATGKVVVMYKVAEQTEFLAL 202
 QY 205 LIMEFVGLI-LPPGVVNVVNGAGVIGEYLATSKRIKVAFTGSTEVGQIQMVA-TONI 262
 Db 203 YVANIKEAGFPFGVNVIPVPGPTAGAAIASHEDVDKVAFTGSTEVGHLITQVAGSNL 262
 QY 263 IPTVLELGKSPNIVFADWDEDAFFDKALE--GFAI FAFNQGVEVTCPSPALVQESIY 320
 Db 263 KRVTLLEGKSPNLS-----DADMDWAVEQAHFALF-FNQSGCCAGSRTVEQIVY 315

QY 421 EFMERAIKRVESIRSNPLDSVTQMTAQVSHQALELILNYHGGKKEGALVLDGKPKL 480
 Db 316 DEFVRSVAPKSPVVGNPDSKTFQGVDFQFKKIIIGYTKSGQGPQAKLLASG 471
 QY 381 LEGFLKD-GYILEPTILFG--QNNMRFVQEFIPGVIAVTEKTEMEALELANDTQYIGAS 447
 Db 372 --CAAAIDRGYFIQPTV FGVAVKICMTIAKELFGVWGLIKKILFEEVVGKANNKYLGA 428
 QY 438 AGVWSRNLAYKMGRTIAGRWNTNYIYAYIAIAAFAGYKSGIGRETHKMKMLHLYQQT 497
 Db 429 AAVTKELDKANYLSQALQAGTVMNCLYDFGCAUSFFGGYKMSGSGRELGEGLQAYTEV 488
 QY 498 KCLLVSYSDK 507
 Db 489 KTVTKVPQK 498

RESULT 13

DHAM_BOVIN ID DHAM_BOVIN STANDARD; PPT; 520 AA.
 AC P20000;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Aldehyde dehydrogenase, mitochondrial precursor (EC 1.2.1.3) (ALDH class 2) (ALDH-E2).
 GN ALDH2.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=90179198; PubMed=1689984;
 RA Guan K., Weiner H.;
 RT "Sequence of the precursor of bovine liver mitochondrial aldehyde dehydrogenase as determined from its cDNA, its gene, and its functionality.";
 RT Arch. Biochem. Biophys. 277:451-460(1990).
 KL [2]
 KN SEQUENCE OF 92-520 FROM N.A.
 PX MEDLINE=89210865; PubMed=2540003;
 RA Farres J., Guan K.L., Weiner H.;
 RT "Primary structures of rat and bovine liver mitochondrial aldehyde dehydrogenases deduced from cDNA sequences.";
 RT Eur. J. Biochem. 180:67-74(1989).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.65 ANGSTROMS).
 PX MEDLINE 97341242; PubMed=9195888;
 RA Steigmetz C.G., Xie P., Weiner H., Burley I.D.;
 RT "Structure of mitochondrial aldehyde dehydrogenase; the genetic component of ethanol aversion.";
 RT Structure 5:701-711(1997).
 CC -|- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
 CC -|- PATHWAY: Ethanol utilization; second step.
 CC -|- SUBUNIT: HOMOTETRAMER.
 CC -|- SUBCELLULAR LOCATION: Mitochondrial matrix.
 CC -|- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 DR PIR: S03565; S03565.
 DR PIR: S04030; S04030.
 DR PIR: IAG8; 08-OCT-97.
 DR FEB, IAG8; 08-APR-98.
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldedh; 1.
 DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; 1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GH; 1.
 KW Oxidoreductase; NAD; Mitochondrion; Transit peptide; 3D structure.
 FT TRANSIT 1 21 MITOCHONDRION.
 FT CHAIN 22 520 ALDEHYDE DEHYDROGENASE.
 FT NP_BIND 265 270 NAD (ADP PART) (BY SIMILARITY).
 FT ACT_SITE 288 288 BY SIMILARITY.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - protein search, using sw model

Run on: June 24, 2003, 10:04:15 ; Search time 39.1863 seconds

(without alignments)
2692.168 Million cell updates/sec

Title: US-09-830-751-8

Perfect score 2673

Sequence: 1 MTNPPSAGIKRQGVGFLK.....HYGKLLVSYSDKPLQLFSLI2

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phase.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_virus.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2651	99.2	542	16 Q8XDJ1	Q8XDJ1 Escherichia
2	2601	97.3	512	16 Q8Z172	Q8Z172 salmonella
3	2589	96.9	512	16 Q8Z295	Q8Z295 salmonella
4	1948	72.9	502	16 Q8Z270	Q8Z270 rhizobium m
5	1918	71.8	506	16 Q8A211	Q8A211 carotobacter
6	1905	71.3	505	16 Q88806	Q88806 rhizobacteri
7	1845	70.4	540	16 Q8UD43	Q8UD43 xanthobacte
8	1879	70.3	505	2 Q50203	Q50203 xanthobacte
9	1878	70.3	506	16 Q8KFL6	Q8KFL6 bacillus ba
10	1870	70.0	506	2 Q8Z918	Q8Z918 petroleum-d
11	1850.5	69.2	503	2 Q8FDS1	Q8FDS1 acinetobact
12	1850.5	69.2	506	2 Q50206	Q50206 xanthobacte
13	1848.5	64.2	505	16 Q8EY70	Q8EY70 bruceella me
14	1802	67.4	506	16 Q8KH05	Q8KH05 pseudomonas
15	1791	67.0	506	16 Q8KR28	Q8KR28 vibrio chol
16	1791	67.0	506	16 Q81254	Q81254 pseudomonas

17	1789	66.9	506	2 Q9ZAA1	Q9ZAA1 pseudomonas
18	1781	66.8	506	2 Q8FHO2	Q8FHO2 carotobacter
19	1780.5	66.6	506	2 Q50009	Q50009 alteromonas s
20	1774	66.4	506	2 Q93Q42	Q93Q42 halomonas s
21	1772	66.3	506	16 Q8XU08	Q8XU08 ralstonia s
22	1756	65.7	506	2 Q8AP43	Q8AP43 yersinia ps
23	1755	65.7	506	2 Q8AG08	Q8AG08 yersinia ps
24	1754	65.6	506	2 Q8AC44	Q8AC44 yersinia ps
25	1257.5	47.0	491	16 Q8R663	Q8R663 fusobacteri
26	1246.5	46.6	495	16 Q8XK44	Q8XK44 staphylococ
27	1052	39.4	499	2 Q8X908	Q8X908 sphingomon
28	1038	38.8	494	2 Q8RK35	Q8RK35 bacillus st
29	1032	38.6	496	16 Q8K112	Q8K112 bacillus ka
30	1021.5	38.2	537	5 Q8TXM0	Q8TXM0 caenorhabdi
31	1010	37.8	501	2 Q85973	Q85973 sphingomon
32	1008.5	37.7	511	2 Q8WXH6	Q8WXH6 alcaligenes
33	986	36.9	495	16 Q34660	Q34660 bacillus su
34	972	36.4	501	2 Q823W8	Q823W8 sphingomon
35	970	36.3	503	3 Q14293	Q14293 scirpocacch
36	969.5	36.3	507	10 Q94JC6	Q94JC6 oryza sativ
37	963.5	36.0	510	5 Q20780	Q20780 caenorhabdi
38	958.5	35.9	494	3 Q8XU14	Q8XU14 neurospora
39	954	35.7	516	13 Q8Z5C2	Q8Z5C2 brachydanio
40	954	35.7	518	13 Q80X03	Q80X03 brachydanio
41	953.5	35.7	511	10 Q8S529	Q8S529 zea mays (m
42	953	35.7	518	13 Q8Q601	Q8Q601 brachydanio
43	952	35.6	518	13 Q80X58	Q80X58 brachydanio
44	952	35.6	519	11 Q8C2S1	Q8C2S1 mus musculu
45	949	35.5	520	5 Q9VLC5	Q9VLC5 drosophila

ALIGNMENTS

RESULT 1

ID	Q8XDJ1	PRELIMINARY:	PRT:	542 AA.
AC	Q8XDJ1;			
DT	01-MAR-2002 (TrEMBLrel. 20, Created)			
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)			
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)			
DE	Aldehyde dehydrogenase B (lactaldehyde dehydrogenase).			
GN	ALDH B 75008 (P. 34464).			
OS	Escherichia coli O157:H7.			
OC	Bacteria; Proteobacteria, gamma subdivision, Enterobacteriaceae,			
OC	Escherichia.			
OX	NCBI_TaxID=83334;			
RN	[1]			
PP	SEQUENCE FROM N.A.			
RC	STRAIN=O157:H7 / EDL933 / ATCC 700927;			
RX	MEDLINE=21074935; PubMed=11206551;			
RA	Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,			
RA	Pose P.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,			
RA	Postfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,			
RA	Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamowski K.,			
RA	Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,			
RA	Welch P.A., Rattner P.P.;			
FT	*Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.;			
RL	Nature 403:529-533(2001).			
EN	[2]			
PP	SEQUENCE FROM N.A.			
RC	STRAIN=O157:H7 / FIMD 0509952;			
RX	MEDLINE=21156231; PubMed=11258796;			
RA	Hayashi T., Makino K., Ohnishi M., Murata T., Tanaka M., Taka T.,			
RA	Han C.-G., Ohnishi E., Nakayama K., Nakayama C., Ogasawara N., Yasunaga T.,			
RA	Iida T., Takami H., Honda T., Sasakawa C., Shinagawa H.,			
RA	Kubara S., Shiba T., Hattori M., Shinagawa H.,			
FT	*Complete genome sequence of enterohaemorrhagic Escherichia coli			
PL	O157:H7 and genomic comparison with a laboratory strain K-12.;			
PL	PNA pos 8:11-22(2001).			
DR	EMBL; AF005585; AAG58731.1;			
DR	EMBL; AF002566; BAB37887.1;			
DP	InterPro; IPR020866; Aldehyde_dehydr			

DR Pfam: PF00171; aldehyd; 1.
 DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
 KW Complete proteome.
 SQ SEQUENCE 542 AA; 59/68 MW; FA08/TA4BEE502A70 CRC64;
 Query Match 99.2%; Score 2651; DB 16; Length 542;
 Best Local Similarity 99.2%; Pred. No. 1.4e-198;
 Matches 508; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MTNPPSAQIKPGEGYFPLKARYDNFTGGWVAFADGEYYONIPTVTGQLLCEVASSG 60
 DL 1 MTNPPSAQIKPGEGYFPLKARYDNFTGGWVAFADGEYYONIPTVTGQLLCEVASSG 90
 QY 61 KRDIIDALDAHKVKQKWAHTSVQDRAAIIFKIADPMQONIFLATAETWNGKPIRETS 120
 DB 91 KRDIIDALDAHKVKQKWAHTSVQDRAAIIFKIADPMQONIFLATAETWNGKPIRETS 150
 QY 121 AADVPLAIDHFRYFASCIKRAQEGGISEVDSSETVAYHFHPLGVGGQIIPWNEPPLMASWK 180
 DB 151 AADVPLAIDHFRYFASCIKRAQEGGISEVDSSETVAYHFHPLGVGGQIIPWNEPPLMASWK 210
 QY 181 MAPALAGNCVVLKPARLTPSLVLLMEIVGDLIPGGVVNVVNGAGGIVGEVLAISKRIA 240
 DB 211 MAPALAGNCVVLKPARLTPSLVLLMEIVGDLIPGGVVNVVNGAGGIVGEVLAISKRIA 270
 QY 241 KVAFTGSTEVGOOIMQYATONIIPTVLELGKSPNIFVADVMDEDAFDKALEGFALFA 300
 DB 271 KVAFTGSTEVGOOIMQYATONIIPTVLELGKSPNIFVADVMDEDAFDKALEGFALFA 330
 QY 301 FNGEVCITCSKALVQESIVERMEKARFVRSFSGNPLISVLMIAQVSHSQLETTLN 360
 DB 331 FNGEVCITCSKALVQESIVERMEKARFVRSFSGNPLISVLMIAQVSHSQLETTLN 390
 QY 361 YIDIGKKGADVLGGRRKLEGELEKGYLLEFGVNNKVPQEEIFGVLAIVTFK 420
 DB 391 YIDIGKKGADVLGGRRKLEGELEKGYLLEFGVNNKVPQEEIFGVLAIVTFK 450
 QY 421 TMEALFLANDTQYGLGAGVWSNGLAYKMGKGIQAGKRVINCYNHAYPAHAAFGYKQS 480
 DB 451 TMEALFLANDTQYGLGAGVWSNGLAYKMGKGIQAGKRVINCYNHAYPAHAAFGYKQS 510
 QY 481 GIGRETHKMMLEHYQOTKCLLVSYSDKPLGLF 512
 DB 511 GIGRETHKMMLEHYQOTKCLLVSYSDKPLGLF 542
 RESULT 2
 Q8ZL72
 ID Q8ZL72 PRELIMINARY; PRT: 512 AA.
 AC Q8ZL72; 2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Aldehyde dehydrogenase B (lactaldehyde dehydrogenase); (EC 1.2.1.22).
 GN ALDB OR STM3680.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae.
 CC Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPAIN-LT2 / SGSC1412 / ATCC 700720.
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Tayman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
 LT2.";
 RT Nature 413:852-856(2001).
 RL

DR EMBL: AE008870; AAL22539.1; ..
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldehyd; 1.
 DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
 KW Oxidoreductase; Complete proteome.
 SQ SEQUENCE 512 AA; 56/126 MW; 770189E45E5GR7DA CRC64;
 Query Match 97.4%; Score 2601; DB 16; Length 512;
 Best Local Similarity 96.9%; Pred. No. 1.1e-194;
 Matches 494; Conservative 11; Mismatches 7; Indels 0; Gaps 0;
 QY 1 MTNPPSAQIKPGEGYFPLKARYDNFTGGWVAFADGEYYONIPTVTGQLLCEVASSG 60
 DB 1 MTNPPSAQIKPGEGYFPLKARYDNFTGGWVAFADGEYYONIPTVTGQLLCEVASSG 60
 QY 61 KRDIIDALDAHKVKQKWAHTSVQDRAAIIFKIADPMQONIFLATAETWNGKPIRETS 120
 DB 61 KRDIIDALDAHKVKQKWAHTSVQDRAAIIFKIADPMQONIFLATAETWNGKPIRETS 120
 QY 121 AADVPLAIDHFRYFASCIKRAQEGGISEVDSSETVAYHFHPLGVGGQIIPWNEPPLMASWK 180
 DB 121 AADVPLAIDHFRYFASCIKRAQEGGISEVDSSETVAYHFHPLGVGGQIIPWNEPPLMASWK 180
 QY 181 MAPALAGNCVVLKPARLTPSLVLLMEIVGDLIPGGVVNVVNGAGGIVGEVLAISKRIA 240
 DB 181 MAPALAGNCVVLKPARLTPSLVLLMEIVGDLIPGGVVNVVNGAGGIVGEVLAISKRIA 240
 QY 241 KVAFTGSTEVGOOIMQYATONIIPTVLELGKSPNIFVADVMDEDAFDKALEGFALFA 300
 DB 241 KVAFTGSTEVGOOIMQYATONIIPTVLELGKSPNIFVADVMDEDAFDKALEGFALFA 300
 QY 301 FNGEVCITCSKALVQESIVERMEKARFVRSFSGNPLISVLMIAQVSHSQLETTLN 360
 DB 301 FNGEVCITCSKALVQESIVERMEKARFVRSFSGNPLISVLMIAQVSHSQLETTLN 360
 QY 361 YIDIGKKGADVLGGRRKLEGELEKGYLLEFGVNNKVPQEEIFGVLAIVTFK 420
 DB 361 YIDIGKKGADVLGGRRKLEGELEKGYLLEFGVNNKVPQEEIFGVLAIVTFK 420
 QY 421 TMEALFLANDTQYGLGAGVWSNGLAYKMGKGIQAGKRVINCYNHAYPAHAAFGYKQS 480
 DB 421 TMEALFLANDTQYGLGAGVWSNGLAYKMGKGIQAGKRVINCYNHAYPAHAAFGYKQS 480
 QY 481 GIGRETHKMMLEHYQOTKCLLVSYSDKPLGLF 512
 DB 481 GIGRETHKMMLEHYQOTKCLLVSYSDKPLGLF 512
 RESULT 3
 Q8ZD05
 ID Q8ZD05 PRELIMINARY; PRT: 512 AA.
 AC Q8ZD05; 2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Aldehyde dehydrogenase B (EC 1.2.1.22).
 GN STY4116.
 OS Salmonella typhi.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Salmonella.
 OX NCBI_TaxID=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPAIN-LT2 / SGSC1412 / ATCC 700720.
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Tayman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
 LT2.";
 RT Nature 413:852-856(2001).
 RL

RA Whitehead S., Barrell B.G.,
 RT "Complete genome sequence of a multiple drug resistant *Salmonella*
 RT *enterica* serovar Typhimurium CT18.",
 RI Nature 413:848-852(2001).
 RL EMBL; AL627281; CAL07945.1;
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldedh; 1.
 DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLD; 1.
 KW Oxidoreductase, Complete proteome.
 SQ SEQUENCE 512 AA; 56355 MW; A64CD3C0FA7E394D CRC64;

Query Match 96.9%; Score 2589; DB 16; Length 512;
 Best Local Similarity 95.7%; Pred. No 9,2e-194;
 Matches 490; Conservative 14; Mismatches 8; Indels 0; Gaps 0;

QY 1 MTNPPSAQIKPGEGEPLKARYDNFIFGEWVAPADGEYQNLTPVTGQLLCEVASSG 60
 DB 1 MTNPPSPTRIQSEYGYPLKARYDNFIFGDNWVAPADGEYQNLTPVTGQPLCEVASSG 60
 QY 61 KRDIPLDALDAAHKVKWAHTSVQDRAAILFKIADRMENLELLATAETWDCPKPIRETS 120
 DB 61 KKDLPLDALDAAHKAKIWAHTSVQDRAAILFKIAIWMFNLELLATAETWDCPKPIRETS 120
 QY 121 AADVPLAIDHFRYFASCIQAEVGGISEVDSSETVAYHFHEPLGVVGQIIPWNPFLLMASWK 180
 DB 121 AADIPLAIDHFRYFASCIQAEVGGISEVDSSETVAYHFHEPLGVVGQIIPWNPFLLMASWK 180
 QY 181 MAPALAGNCVVLKPARLITPLSVLLLMFIVGDLPLPGGVVNVNAGGVIGEYLATSKRIA 240
 DB 181 MAPALAGNCVVLKPARLITPLSVLLLMFIVGDLPLPGGVVNVNAGGVIGEYLATSKRIA 240
 QY 241 KVAFTGSTVEVGGQIMQVATONIIPVTLFGLGKSPNIFADVMDDEDAFFDKALEGFALFA 300
 DB 241 KVAFTGSTVEVGGQIMQVATONIIPVTLFGLGKSPNIFADVMDDEDAFFDKALEGFALFA 300
 QY 301 FNOGEVCTCPSRALVQRSIYEPFNERAIPRVESIPSONPLDSVTQMGAVSHQLETLIN 360
 DB 301 FNOGEVCTCPSRALVQRSIYEPFNERAIPRVESIPSONPLDSVTQMGAVSHQLETLIN 360
 QY 361 YIDICKKGGANVLTGGRRKILAFGEUKDGYLYLEPTIFPGNNMRVFEIEFGPVLAVTTFK 420
 DB 361 YIDICKKGGADLTGGRRKELDGEIKGYLYLEPTILFGKNNMRVFEIEFGPVLAVTTFK 420
 QY 421 TMEALELANDIQGLGAGVWSRNGNIAYKMGRIQAGRVNTNCYHAYPAHAAFGGYKQS 480
 DB 421 TMEKALELANDIQGLGAGVWSRNGNIAYKMGRIQAGRVNTNCYHAYPAHAAFGGYKQS 480
 QY 481 GIGRETHKMMLEHYOQTKCLLVSYSDKPLGLF 512
 DB 481 GIGRETHKMMLEHYOQTKCLLVSYSDKPLGLF 512

RESULT 4

Q92N70
 ID Q92N70 PRELIMINARY; PRT; 502 AA.
 AC Q92N70;
 DT 01-DEC-2001 (TRENBLrel. 19, Created)
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE Probable aldehyde dehydrogenase protein (EC 1.2.1.3).
 GN R02351 (K. SWC02689).
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396507; PubMed=11481430;
 RA Capella D., Barloy-Hubler F., Gouzy T., Rothe G., Ampe F., Batut J.,
 RA Boistard P., Becker A., Bouthry M., Cadieu F., Dreano S., Gloux S.,
 RA Godrie T., Goffeau A., Kahn D., Kiss F., Lelaure V., Masuy D.,

RESULT 5

Q9A211
 ID Q9A211 PRELIMINARY; PRT; 506 AA.
 AC Q9A211;
 DT 01-JUN-2001 (TRENBLrel. 17, Created)
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
 DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
 DE Aldehyde dehydrogenase B.
 GN CC3580.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
 OC Caulobacter.
 OX NCBI_TaxID=155892;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 19089 / CB15;
 RX MEDLINE=21173698; PubMed=11259647;
 RA Nierman W.C., Feldhym T.V., Laub M.T., Paulsen I.T., Nelson K.F.,
 RA Eisen J., Heidelberg J.F., Alley M.B.K., Ohta N., Muddock J.P.,

BA Pohl T., Poutetelle D., Puchler A., Poutetelle B., Ramsperger U.,
 BA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.,
 RI "Analysis of the chromosome sequence of the legume symbiont
 RI *Sinorhizobium meliloti* strain 1021.",
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 DR EMBL; AL591790; CAC46930.1;
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldedh; 1.
 DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLD; UNKNOWN_1.
 KW Oxidoreductase; Complete proteome.
 SQ SEQUENCE 502 AA; 55250 MW; D45A262260REB436 CRC64;

Query Match 72.9%; Score 1948; DB 16; Length 502;
 Best Local Similarity 73.5%; Pred. No 1.1e-143;
 Matches 361; Conservative 51; Mismatches 79; Indels 0; Gaps 0;

QY 22 KARYDNFIFGEWVAPADGEYQNLTPVTGQLLCEVASSGKPIDLALDAAHKVKWAHT 81
 DB 12 KARYDNFIFGEWVAPADGEYQNLTPVTGQLLCEVASSGKPIDLALDAAHKVKWAHT 81
 QY 82 SVQDRAAILFKIADRMENLELLATAETWDCPKPIRETS 141
 DB 72 SVQDRAAILFKIADRMENLELLATAETWDCPKPIRETS 141
 QY 142 EGGISEVDSSETVAYHFHEPLGVVGQIIPWNPFLLMASWKMAPALAGNCVVLKPARLITPL 201
 DB 132 EGGISEVDSSETVAYHFHEPLGVVGQIIPWNPFLLMASWKMAPALAGNCVVLKPARLITPL 201
 QY 202 SVLLMEIVGDLPLPGGVVNVNAGGVIGEYLATSKRIAFTSTVEVGGQIMQVATON 261
 DB 192 SVLLMEIVGDLPLPGGVVNVNAGGVIGEYLATSKRIAFTSTVEVGGQIMQVATON 261
 QY 262 IIPVTLFGLGKSPNIFADVMDDEDAFFDKALEGFALFANOGEVCTCPSRALVQRSIYE 321
 DB 252 IIPVTLFGLGKSPNIFADVMDDEDAFFDKALEGFALFANOGEVCTCPSRALVQRSIYE 321
 QY 322 REMERAIIPVESIPSONPLDSVTQMGAVSHQLETLINIDTICKKPGVAVTTPK 381
 DB 312 REMERAIIPVESIPSONPLDSVTQMGAVSHQLETLINIDTICKKPGVAVTTPK 381
 QY 382 FRELKIKYLYLEPTILPGNNMRVFEIEFGPVLAVTTFK 441
 DB 372 FRELKIKYLYLEPTILPGNNMRVFEIEFGPVLAVTTFK 441
 QY 442 SNGNLAIVKMGRIQAGRVNTNCYHAYPAHAAFGGYKQSIFPFHFMMLHYOQTKCL 501
 DB 432 SNGNLAIVKMGRIQAGRVNTNCYHAYPAHAAFGGYKQSIFPFHFMMLHYOQTKCL 501
 QY 502 VSYSDKPLGLF 512
 DB 492 VSYSDKPLGLF 502

RA Potocka L., Nelson W.C., Newton A., Stephens C., Phadke N.D., Fly R.,
 RA DeBoy R.T., Dodson P.J., Durkin A.S., Gwinn M.L., Haft D.H.,
 RA Kolonay J.F., Smir J., Graven M.R., Khouri H., Sherry J., Berry K.,
 RA Utterback T., Tran K., Wolf A., Vamathevan J., Frimolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.,
 RT "Complete genome sequence of *Caulobacter crescentus*,"
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 DR EMBL: AF060116; AAK25542.1; ..
 DR HSP: P05091; ICW3.
 DR TIGR: CC3580; ..
 DR InterPro: IP0002086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldedh; 1.
 DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
 KW Complete proteome.
 SQ SEQUENCE 506 AA: 55661 MW: 268909ALE52600 CRG64.

Query Match 71.8%; Score 1918; DB 16; Length 506;

Best Local Similarity 71.7%; Pred. No. 2.4e-141;

Matches 352; Conservative 57; Mismatches 82; Indels 0; Gaps 0;

QY 22 KARYDNFEGEWAPADGEYQNLTPVTGOLCEVASSCKRDLALDAAHKVKKRAHT 81
 DB 16 KARYDNFEGEWAPADGEYQNLTPVTGOLCEVASSCKRDLALDAAHKVKKRAHT 75
 QY 82 SVQDRAAILEKIAADRMFONLELLATAETWNGKPIRETSAADVPLAIDHFRYFASCIRAO 141
 DB 76 SAADRSILLRIADRMFONLELLATAETWNGKPIRETSAADVPLAIDHFRYFASCIRAO 135
 QY 142 EGGISEVDSTVAYHHEPLGVVGQIIPWNPFLMASKMAPALAGNCVVLKPARLTP. 201
 DB 136 EGGISEVDSTVAYHHEPLGVVGQIIPWNPFLMASKMAPALAGNCVVLKPARLTP. 195
 QY 202 SVLLMEIVGDLPPGVVNVVNGAGGVEIGYLAETSKRIAKVAFSGSIEVGQIIMYATON 261
 DB 196 SIMWAEIMIGDLPPGVVNVVNGAGGVEIGYLAETSKRIAKVAFSGSIEVGQIIMYATON 255
 QY 262 IIPVTLELGKSNIPVADWDEDAFFDKALEGALFAGNVEGCTCPSRALQVESIVE 321
 DB 256 IIPVTLELGKSNIPVADWDEDAFFDKALEGALFAGNVEGCTCPSRALQVESIVE 315
 QY 322 KFERALRRVESIRSGNPLDSVTOMGAVSHQLETLINNYIDIGKEGADVLTGRRKLL 381
 DB 316 KFERALRRVESIRSGNPLDSVTOMGAVSHQLETLINNYIDIGKEGADVLTGRRKLL 375
 QY 382 EGELKDCGYLLEPTILFGQNNMRFQEEIFGPVLAVTTTEKMEALELANDTOYGLGAGW 441
 DB 376 PGQLADGYVEPTVFECHNMRIFQEEIFGPVLAVTTTEKMEALELANDTOYGLGAGW 435
 QY 442 SRNCNLAJMGRCIOACRWNTNCHAYPAHAFAFGYKQSGIGRETHKMMLEHYQOTKLL 501
 DB 436 SRNCNLAJMGRCIOACRWNTNCHAYPAHAFAFGYKQSGIGRETHKMMLEHYQOTKLL 495
 QY 502 VSYSDKPIGLF 512
 DB 496 VSYSPRALGPF 506

RESULT 6

Q98806
 ID Q98806 PRELIMINARY; PRT; 505 AA.
 AC Q98806;
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, last annotation update)
 DE Aldehyde dehydrogenase
 GN ML6639.
 OS *Mesorhizobium loti* (*Mesorhizobium loti*)
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID:381;
 RN [1]
 RP SEQUENCE FROM N.A.

PC STRAIN=MAFF303099;
 RX MEDLINE=21082940; PubMed 11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu F., Kato T., Sasamoto S.,
 RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kikuchi T.,
 RA Kishida Y., Kiyokawa G., Kohata M., Matsumoto M., Matsuno A.,
 RA Morishizuki Y., Nakayama S., Nakazaki N., Shimizu S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.,
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti,"
 RL DNA Res. 7:331-338(2000).
 DR EMBL: AP004009; BAB52891.1; ..
 DR InterPro: IP0002086; Aldehyde_dehydr.
 DR Pfam: PF00171; aldedh; 1.
 DR PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
 DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
 KW Complete proteome.
 SQ SEQUENCE 505 AA: 55586 MW: 158763AFAH0FA37 CRG64;

Query Match 71.3%; Score 1905; DB 16; Length 505;
 Best Local Similarity 70.9%; Pred. No. 2.5e-140;
 Matches 352; Conservative 63; Mismatches 84; Indels 2; Gaps 1;

QY 14 EYGFPLK--LKARYDNFEGEWAPADGEYQNLTPVTGOLCEVASSCKRDLALDA 71
 DB 14 EYGFPLK--LKARYDNFEGEWAPADGEYQNLTPVTGOLCEVASSCKRDLALDA 64
 QY 72 HKYDKWAHTSVQDRAAILEKIAADRMFONLELLATAETWNGKPIRETSAADVPLAIDHFR 141
 DB 65 HAAKDAWGTSTVAFSPHLEPLADRMFONLELLATAETWNGKPIRETSAADVPLAIDHFR 124
 QY 132 RYFASCIROEGGISEVDSTVAYHHEPLGVVGQIIPWNPFLMASKMAPALAGNCV 191
 DB 125 RYFASCIROEGGISEVDSTVAYHHEPLGVVGQIIPWNPFLMASKMAPALAGNCV 184
 QY 192 VFPAPLIPPLSVLLMEIVGDLPPGVVNVVNGAGGVEIGYLAETSKRIAKVAFSGSIEVG 251
 DB 185 VFPAPLIPPLSVLLMEIVGDLPPGVVNVVNGAGGVEIGYLAETSKRIAKVAFSGSIEVG 244
 QY 252 QCLMAYATONIIIPVTLELGKSNIPVADWDEDAFFDKALEGALFAGNVEGCTCPS 411
 DB 245 QCLMAYATONIIIPVTLELGKSNIPVADWDEDAFFDKALEGALFAGNVEGCTCPS 404
 QY 312 PAIVQESTIVERPMEPAIPVFSIPSCNPLDSVTOMGAVSHQLETLINNYIDIGKEGADV 471
 DB 305 PAIVQESTIVERPMEPAIPVFSIPSCNPLDSVTOMGAVSHQLETLINNYIDIGKEGADV 464
 QY 372 VLTGGRKRLLEGKDCGYLLEPTILFGQNNMRFQEEIFGPVLAVTTTEKMEALELAND 431
 DB 365 VLTGGRKRLLEGKDCGYLLEPTILFGQNNMRFQEEIFGPVLAVTTTEKMEALELAND 424
 QY 432 TOYGLGAGWVSRNCLAYMGRCIOACRWNTNCHAYPAHAFAFGYKQSGIGRETHKMM 491
 DB 425 TOYGLGAGWVSRNCLAYMGRCIOACRWNTNCHAYPAHAFAFGYKQSGIGRETHKMM 484
 QY 492 EHYQOTKLLHYQOTKLL 512
 DB 485 EHYQOTKLLHYQOTKLL 505

RESULT 7

Q98806
 ID Q98806 PRELIMINARY; PRT; 505 AA.
 AC Q98806;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, last annotation update)
 DE Aldehyde dehydrogenase
 GN AL6A.GR AT02224 GR ACB_C_4041.
 OS *Acetivibrio tumefaciens* (strain C5b / ATCC 34970).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID:176299;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=21608550; PubMed=11743193;
 PA Wood P.W., Serfaty T.C., Kaul P., Monks P.F., Kitajima T.P.,
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyce D.S.,
 RA Chapman P., Clendenning J., Weatherage G., Gillet W., Grant C.,
 PA Kutayvin T., Levy R., Li M.-J., McLelland E., Palmeri A.,
 RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
 RA Zhang S., Yoo H., Tan Y., Biddle P., Jung M., Krespan W., Perry M.,
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Nestor E.W.,
 RT "The genome of the natural genetic engineer *Agrobacterium tumefaciens*
 PT C58."
 PL Science 294:2317-2323(2001).
 RN [12]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21608551; PubMed=11743194;
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
 RA Qurollo B., Goldman B.S., Fan Y., Askenazi M., Halling C., Mullin L.,
 RA Roumel K., Gordon J., Vaudin M., Tarrachou O., Epp A., Liu F.,
 RA Wollam C., Allender M., Doughty D., Scott C., Lappas C., Markelz B.,
 RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
 RA Cieto C., Slater S.;
 RT "Genome sequence of the plant pathogen and biotechnology agent
 PT *Agrobacterium tumefaciens* C58."
 PL Science 294:2317-2323(2001).
 DR EMBL: AF009172; AAL43213.1; ALT_INIT.
 DR FMBL: AF008138; AAK47966.1;
 KW Complete proteome.
 SQ SEQUENCE 540 AA; 59646 MW; B03074B38B1A8E46 CRC64;

Query Match 70.9%; Score 1895; DB 16; Length 540;
 Best Local Similarity 69.0%; Pred. No. 1.7e-139;
 Matches 348; Conservative 68; Mismatches 84; Indels 4; Gaps 1;

QY 9 QIKPGYGFPLKAPYDNFPGGFWAPADGYYQNI:TPVTGQLLCEVASSKRPIDLLAL 68
 DB 41 QOKAGEAPFLK---YNYIGGKWEKSPGYMDNI:SPVTGKIK:FPVPSDASDTEPAL 96
 QY 69 DAHKKVKQWHTSVQDPAALFKIADPMEONLELLATAETWNGKPIRPIETSAADVPLAI 128
 DB 97 DAHKKVKQWHTSVQDPAALFKIADPMEONLELLATAETWNGKPIRPIETSAADVPLAI 156
 QY 129 DFRYFASCIRAQEGGISEVDSETVAYHFHEPLGVGQIIPWNPFLLMASWKPALAAAG 188
 DB 157 DFRYFASCIRAQEGGISEVDSETVAYHFHEPLGVGQIIPWNPFLLMASWKPALAAAG 216
 QY 189 NPVVIKPAELIPLSVILLMEIVGILIPGVVNVVNGASGVIGEYLAISKIAVFTGST 248
 DB 217 NPVVIKPAELIPLSVILLMEIVGILIPGVVNVVNGASGVIGEYLAISKIAVFTGST 276
 QY 249 EYGGQIMGYATQNIIPVTLELGGKSPNIVFALVMEFALFUKALEGFALFAPNAGEVCT 308
 DB 277 SVGKELMYADNVTNISLELGGKSPNIVFADVWNEUADFLKALEGFALFAPNAGEVCT 336
 QY 309 CPSRALVQESYVERPMEPATRVPSTFSGNPLDSVTOMGACVSHGQLETLINIDIGKKE 368
 DB 337 CPSRALVQESYVERPMEPATRVPSTFSGNPLDSVTOMGACVSHGQLETLINIDIGKKE 396
 QY 369 GATVIGGFKLLELEKLYYIEPTILFGNNMPVLEPFPSPVLAVTFKTVMEALEL 428
 DB 397 GATVIGGFKLLELEKLYYIEPTILFGNNMPVLEPFPSPVLAVTFKTVMEALEL 456
 QY 429 ANDUYGLGAGVWSKNGNLAYKMGKGLQAGRWNTNCHAYPAHAAGFYKQSGIGRETHK 488
 DB 457 ANDUYGLGAGVWSKNGNLAYKMGKGLQAGRWNTNCHAYPAHAAGFYKQSGIGRETHK 516
 QY 489 MMEHYQOTKCLLVSYSDKPIGLF 512
 DB 517 MMEHYQOTKCLLVSYSDKPIGLF 540

RESULT 8
 Q56203 PPEIIMINAPY; PPT; 505 AA
 ID Q56203;
 DT 01-JUN-1998 (Tremblrel. 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Chloroacetaldehyde dehydrogenase-
 GN ALDA.
 OS Xanthobacter autotrophicus
 OS Plasmid linear plasmid pXA01.
 OC Bacteria, Proteobacteria; alpha subdivision, Rhizobiaceae group;
 OC Hyphomicrobium group; Xanthobacter.
 OX NCBI_TaxID=280;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98172729; PubMed=9511738;
 RA Bergeron H., Iatbe D., Turmel C., Lau P.C.;
 RT "Cloning, sequence and expression of a linear plasmid based and a
 RT chromosomal homolog of chloroacetaldehyde dehydrogenase-encoding genes
 RT in Xanthobacter autotrophicus GJ10."
 RL Gene 207:9-18(1998).
 CC -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
 DF FMBL: AF029733; AAC13641.1;
 DF HSP: P05091; IOW3.
 DF IUB: P05091; IUB002386; A3dehyd_dehyd;
 DF Pfam: PF00171; aldehyd; 1.
 DF PROSITE: PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1
 DF PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
 KW Oxidoreductase; Plasmid.
 SQ SEQUENCE 505 AA; 54945 MW; AD6A9EDB0297BB04 CRC64;

Query Match 70.3%; Score 1879; DB 2; Length 505;
 Best Local Similarity 69.2%; Pred. No. 2.7e-138;
 Matches 353; Conservative 58; Mismatches 93; Indels 6; Gaps 1;

QY 3 NNPI:SAQIKPGYGFPLKAPYDNFPGGFWAPADGYYQNI:TPVTGQLLCEVASSKRP 62
 DB 2 NKPEIATIKQSPF-----KAPYGNFTGGKVEPVGPGPYFNT:SPVTGKIK:ICIAFSAD 55
 QY 63 DIDALDAHKKVKQWHTSVQDPAALFKIADPMEONLELLATAETWNGKPIRPIETSA 122
 DB 56 DIEKALDAHAAKAWGKTSAAVRALILNKAIDRMEENLDLLALAEITWNGKPIRPIET 115
 QY 123 DVPALDAIFRYFASCIRAQEGGISEVDSETVAYHFHEPLGVGQIIPWNPFLLMASW 182
 DB 116 DVPALDAIFRYFASCIRAQEGGISEVDSETVAYHFHEPLGVGQIIPWNPFLLMASW 175
 QY 183 DALAAGNPVVIKPAELIPLSVILLMEIVGILIPGVVNVVNGASGVIGEYLAISKIAV 242
 DB 176 DALAAGNPVVIKPAELIPLSVILLMEIVGILIPGVVNVVNGASGVIGEYLAISKIAV 245
 QY 243 APTGCTFVGGQIMGYATQNIIPVTLELGGKSPNIVFALVMEFALFUKALEGFALFAPN 402
 DB 236 APTGCTFVGGQIMGYATQNIIPVTLELGGKSPNIVFADVWNEUADFLKALEGFALFAPN 295
 QY 303 QSEVTEPSRALVQESYVERPMEPATRVPSTFSGNPLDSVTOMGACVSHGQLETLIN 362
 DB 296 QSEVTEPSRALVQESYVERPMEPATRVPSTFSGNPLDSVTOMGACVSHGQLETLIN 355
 QY 363 GATVIGGFKLLELEKLYYIEPTILFGNNMPVLEPFPSPVLAVTFKTVMEALEL 422
 DB 356 GATVIGGFKLLELEKLYYIEPTILFGNNMPVLEPFPSPVLAVTFKTVMEALEL 415
 QY 423 FEALFLANTUYGLGAGVWSKNGNLAYKMGKGLQAGRWNTNCHAYPAHAAGFYKQSGI 492
 DB 416 FEALFLANTUYGLGAGVWSKNGNLAYKMGKGLQAGRWNTNCHAYPAHAAGFYKQSGI 475
 QY 483 GPETHKMMLEHYQOTKCLLVSYSDKPIGLF 512
 DB 476 GPETHKMMLEHYQOTKCLLVSYSDKPIGLF 505

